INSTITUTE FOR ADVANCED STUDIES

JODHPUR

CSIR NET LIFESCIENCES

STUDY MATERIAL

SET-1

UNIT: 1-3

INSTITUTE FOR ADVANCED STUDIES
B-7 SARASWATI NAGAR, JODHPUR (RAJ.)
0291-2721056
e-mail:ifasnet@gmail.com
http://csirnetlifesciencs.tripod.com

BLANK PAGE

INSTITUTE FOR ADVANCED STUDIES (IFAS)

B-7 SARASWATI NAGAR, BASNI-I, JODHPUR (RAJ)

Contact: 0291-2721056; 09460660533

e-mail:ifasnet@gmail.com

http://csirnetlifesciences.tripod.com

IFAS

JODHPUR

CONTENTS

UNIT-1

1. MOLECULES AND THEIR INTERACTION RELAVENT TO BIOLOGY

A. Structure of atoms, molecules and chemical bonds.
B. Composition, structure and function of biomolecules (carbohydrates, lipids, proteins, nucleic acids and vitamins).
C. Stablizing interactions (Van der Waals, electrostatic, hydrogen bonding, hydrophobic interaction, etc.).
D. Principles of biophysical chemistry (pH, buffer, reaction kinetics, thermodynamics, colligative properties).
E. Bioenergetics, glycolysis, oxidative phosphorylation, coupled reaction, group transfer, biological energy transducers.
F. Principles of catalysis, enzymes and enzyme kinetics, enzyme regulation, mechanism of enzyme catalysis, isozymes.
G. Conformation of proteins (Ramachandran plot, secondary, tertiary and quaternary structure; domains; motif and folds).
H. Conformation of nucleic acids (A-, B-, Z-,DNA), t-RNA, micro-RNA).
I. Stability of protein and nucleic acid structures.
J. Metabolism of carbohydrates, lipids, amino acids, nucleotides and vitamins.

UNIT-2

2. CELLULAR ORGANIZATION

A. Membrane structure and function: Structure of model membrane, lipid bilayer and membrane protein diffusion, osmosis, ion channels, active transport, ion pumps, mechanism of sorting and regulation of intracellular transport, electrical properties of membranes.
B. Structural organization and function of intracellular organelles: Cell wall, nucleus, mitochondria, Golgi bodies, lysosomes, endoplasmic reticulum, peroxisomes, plastids, vacuoles, chloroplast, structure & function of cytoskeleton and its role in motility.
C. Organization of genes and chromosomes: Operon, interrupted genes, gene families, structure of chromatin and chromosomes, unique and repetitive DNA, heterochromatin, euchromatin, transposons.
D. Cell division and cell cycle: Mitosis and meiosis, their regulation, steps in cell cycle, and control of cell cycle.
E. Microbial Physiology: Growth, yield and characteristics, strategies of cell division, stress response.

UNIT-3

3. FUNDAMENTAL PROCESSES

A. DNA replication, repair and recombination: Unit of replication, enzymes involved, replication origin and replication fork, fidelity of replication, extrachromosomal replicons, DNA damage and repair mechanisms.
B. RNA synthesis and processing: Transcription factors and machinery, formation of initiation complex, transcription activators and repressors, RNA polymerases, capping, elongation and termination, RNA processing, RNA editing, splicing, polyadenylation, structure and function of different types of RNA, RNA transport.
C. Protein synthesis and processing: Ribosome, formation of initiation complex, initiation factors and their regulation, elongation and elongation factors, termination, genetic code, aminoacylation of tRNA, tRNA-identity, aminoacyl tRNA synthetase, translational proof-reading, translational inhibitors, post- translational modification of proteins.
D. Control of gene expression at transcription and translation level: Regulation of phages, viruses, prokaryotic and eukaryotic gene expression, role of chromatin in regulating gene expression and gene silencing.

UNIT-I

MOLECULES AND THEIR INTERACTION RELAVENT TO BIOLOGY

INSTITUTE FOR ADVANCED STUDIES

B-7 SARASWATI NAGAR, JODHPUR

e-mail:ifasnet@gmail.com

http://csirnetlifesciences.tripod.com

BLANK PAGE

A1. Structure of Atoms, Molecules

1. Elements consist of only one kind of atom and cannot be decomposed into simpler substances. Our planet is made up of some 90 elements. (Tiny amounts — sometimes only a few atoms — of additional elements have been made in nuclear physics laboratories, but they play no role in our story). Of these 90, only 25 or so are used to build living things.

2. The table shows the 11 most prevalent elements in the lithosphere (the earth's crust) and in the human body.

3. Living matter uses only a fraction of the elements available to it but, as the table shows, the relative proportions of those it does acquire from its surroundings are quite different from the proportions in the environment.

So, the composition of living things is not simply a reflection of the elements available to them

For example, hydrogen, carbon, and nitrogen together

Elemental composition of the lithosphere and the human body. Each number represents the percent of the total number of atoms present. For example, 47 of every 100 atoms found in a representative sample of the lithosphere are oxygen while there are only 19 atoms of carbon in every 10,000 atoms of lithosphere.

Composition of the Lithosphere		Composition of the Human Body	
Oxygen	47	Hydrogen	63
Silicon	28	Oxygen	25.5
Aluminum	7.9	Carbon	9.5
Iron	4.5	Nitrogen	1.4
Calcium	3.5	Calcium	0.31
Sodium	2.5	Phosphorus	0.22
Potassium	2.5	Chlorine	0.03
Magnesium	2.2	Potassium	0.06
Titanium	0.46	Sulfur	0.05
Hydrogen	0.22	Sodium	0.03
Carbon	0.19	Magnesium	0.01
All others	<0.1	All others	<0.01

represent less than 1% of the atoms found in the earth's crust but some 74% of the atoms in living matter. One of the properties of life is to take up certain elements that are scarce in the nonliving world and concentrate them within living cells.

Some sea animals accumulate elements like vanadium and iodine within their cells to concentrations a thousand or more times as great as in the surrounding sea water. It has even been proposed that uranium be "mined" from the sea by extracting it from certain algae that can take up uranium from sea water and concentrate it within their cells.

4. There is still some uncertainty about the exact number of elements required by living things. Some elements, e.g., selenium and aluminum, are found in tiny amounts within living things, but whether they are playing an essential role (selenium is) or are simply an accidental acquisition (aluminum probably) is sometimes difficult to determine.

5. Atoms: Each element is made up of one kind of atom. We can define an atom as the smallest part of an element that can enter into combination with other elements. Each atom consists of a small, dense, positively-charged nucleus surrounded by much lighter, negatively-charged electrons.

6. The nucleus of the simplest atom, the hydrogen atom (H), consists of a single positively-charged proton. Because of its single proton, the atom of hydrogen is assigned an atomic number of 1 and a single electron.

7. The charge of the electron is the same magnitude as that of the proton, so the atom as a whole is electrically neutral. Its proton accounts for almost all the weight of the atom.

9. The nucleus of the atom of the element helium (He) has two protons (hence helium has an atomic number of 2) and two neutrons. Neutrons have the same weight as protons but no electrical charge. The helium atom has two electrons so that, once again, the atom as a whole is neutral.

10. The structure of each of the other kinds of atoms follows the same plan. From Lithium (At. No. = 3) to uranium (At. No. = 92), the atoms of each element can be listed in order of increasing atomic number. There are no gaps in the list. Each element has a unique atomic number and its atoms have one more proton and one more electron than the atoms of the element that precedes it in the list.

Electrons

1. Electrons are confined to relatively discrete regions around the nucleus. The two electrons of helium, for example, are confined to a spherical zone surrounding the nucleus called the K shell or K energy level.

2. Lithium (At. No. = 3) has three electrons, two in the K shell and one located farther from the nucleus in the L shell. Being farther away from the opposite (+) charges of the nucleus, this third electron is held less tightly.

3. Each of the following elements, in order of increasing atomic number, adds one more electron to the L shell until we reach neon (At. No. = 10) which has eight electrons in the L shell.

4. Sodium places its eleventh electron in a still higher energy level, the M shell.

5. From sodium to argon, this shell is gradually filled with electrons until, once again, a maximum of eight is reached. Note that after the K shell with its maximum of two electrons, the maximum number of electrons in any other outermost shell is eight. As we shall see, the chemical properties of each element are strongly influenced by the number of electrons in its outermost energy level (shell).

6. This table shows the electronic structure of the atoms of elements 1 – 36 with those that have been demonstrated to be used by living things. Four elements of still higher atomic numbers that have been shown to be used by living things are also included.

7. The electronic structure of an atom plays the major role in its chemistry. The pattern of electrons in an atom — especially those in the outermost shell — determines the valence of the atom; that is, the ratios in which it interacts with other atoms, and to a large degree, the electronegativity of the atom; that is, the strength with which it attracts other electrons.

8. Elements with the same number of electrons in their outermost shell show similar chemical properties.

Example 1: Fluorine, chlorine, bromine, and iodine each have 7 electrons in their outermost shell. These so-called halogens are also quite similar in their chemical behavior. When dissolved in water, for example, they all produce germicidal solutions.

Atomic Number	Element	Energy Levels or "shells"				
		K	L	M	N	O
1	Hydrogen (H)	1				
2	Helium (He)	2				
3	Lithium (Li)	2	1			
4	Beryllium (Be)	2	2			
5	Boron (B)	2	3			
6	Carbon (C)	2	4			
7	Nitrogen (N)	2	5			
8	Oxygen (O)	2	6			
9	Fluorine (F)	2	7			
10	Neon (Ne)	2	8			
11	Sodium (Na)	2	8	1		
12	Magnesium (Mg)	2	8	2		
13	Aluminum (Al)	2	8	3		
14	Silicon (Si)	2	8	4		
15	Phosphorus (P)	2	8	5		
16	Sulfur (S)	2	8	6		
17	Chlorine (Cl)	2	8	7		
18	Argon (Ar)	2	8	8		
19	Potassium (K)	2	8	8	1	
20	Calcium (Ca)	2	8	8	2	
21	Scandium (Sc)	2	8	9	2	
22	Titanium (Ti)	2	8	10	2	
23	Vanadium (V)	2	8	11	2	
24	Chromium (Cr)	2	8	13	1	
25	Manganese (Mn)	2	8	13	2	
26	Iron (Fe)	2	8	14	2	
27	Cobalt (Co)	2	8	15	2	
28	Nickel (Ni)	2	8	16	2	
29	Copper (Cu)	2	8	18	1	
30	Zinc (Zn)	2	8	18	2	
31	Gallium (Ga)	2	8	18	3	
32	Germanium (Ge)	2	8	18	4	
33	Arsenic (As)	2	8	18	5	
34	Selenium (Se)	2	8	18	6	
35	Bromine (Br)	2	8	18	7	
36	Krypton (Kr)	2	8	18	8	
42	Molybdenum (Mo)	2	8	18	13	1
48	Cadmium (Cd)	2	8	18	18	2
50	Tin (Sn)	2	8	18	18	4
53	Iodine (I)	2	8	18	18	7

Example 2: Those elements with 1, 2, or 3 electrons in their outermost shell are the metals.
Example 3: Those elements with 4, 5, 6, or 7 in their outermost shell are the nonmetals.

Example 4: Helium (with its 2), neon, argon, and krypton (each with 8) have "filled" their outermost shells. They are the so-called inert or "noble" gases. They have no chemistry at all. Under normal conditions they do not interact with other atoms. So, it is the number and arrangement of the electrons in the atoms of an element that establish the chemical behavior of that element.

9. The atoms of an element interact with other atoms in such ways and ratios that they can "fill" their outermost shell with 8 electrons (2 for hydrogen). They may do this by **acquiring more electrons from another atom, losing electrons to another atom, sharing electrons with another atom**. The number of electrons that an atom must acquire, or lose, or share to reach a stable configuration of 8 (2 for hydrogen) is called its **valence**.

10. Hydrogen, lithium, sodium, and potassium atoms all have a single electron in their outermost shell. Fluorine, chlorine, bromine, and iodine atoms all have 7. Any atom of the first group will interact with a single atom of any of the second group forming, HCl, NaCl, KI, etc. The result of all of these interactions is a pair of atoms each with an outermost shell like that of one of the inert gases: 2 for hydrogen, 8 for the others.

- The elements with 2 electrons in their outermost shell interact with chlorine and the other halogens to form, e.g., $BeCl_2$, $MgCl_2$, $CaCl_2$. Again, the result is a pair of atoms each with a stable octet of electrons in its outermost shell.
- The elements with 3 electrons in their outermost shell will interact with chlorine in a ratio of 1:3, forming BCl_3, $AlCl_3$.
- Carbon atoms, with their 4 electrons in the L shell interact with chlorine to form CCl_4.
- Nitrogen, with its 5 outermost electrons, interacts with hydrogen atoms in a ratio of 1:3, forming ammonia (NH_3).
- Oxygen and sulfur, with their 6 outermost electrons react with hydrogen to form water (H_2O) and hydrogen sulfide (H_2S).

11. What determines whether a pair of atoms swaps or share electrons? The answer is their relative electronegativities. If two atoms differ greatly in their affinity for electrons; that is, in their electronegativity, then the strongly electronegative atom will take the electron away from the weakly electronegative one.

Example 1: Na (weakly electronegative) gives up its single electron to an atom of chlorine (strongly electronegative) to form NaCl. The sodium atom now has only 10 electrons but still 11 protons so there is a net positive charge of one on the atom. Similarly, chlorine now has one more electron than proton so its now has a net negative charge of 1. Electrically charged atoms are called ions. The mutual attraction of opposite electrical charges holds the ions together by ionic bonds.

Example 2: Carbon and hydrogen are both only weakly electronegative so neither can remove electrons from the other. Instead they achieve a stable configuration by sharing their outermost electrons forming covalent bonds of CH4.

Isotopes

1. The number of protons in the nucleus of its atoms, which is its atomic number, defines each element. However, the nuclei of a given element may have varying numbers of neutrons. Because neutrons have weight (about the same as that of protons), such atoms differ in the atomic weight. Atoms of the same element that differ in their atomic weight are called isotopes.

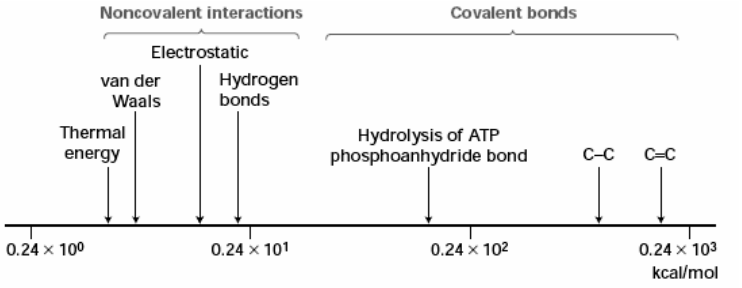
2. Atomic weights are expressed in terms of a standard atom: the isotope of carbon that has 6 protons and 6 neutrons in its nucleus. This atom is designated carbon-12 or 12C. It is arbitrarily assigned an atomic weight of 12 daltons (named after John Dalton, the pioneer in the study of atomic weights). Thus a dalton is 1/12 the weight of an atom of 12C. Both protons and neutrons have weights very close to 1 dalton each. Carbon-12 is the commonest isotope of carbon. Carbon-13 (13C) with 6 protons and 7 neutrons, and carbon-14 (14C) with 6 protons and 8 neutrons are found in much smaller quantities.

3. Isotopes as "tracers": One can prepare, for example, a carbon compound used by living things that has many of its normal ^{12}C atoms replaced by ^{14}C atoms. Carbon-14 happens to be radioactive. By tracing the fate of radioactivity within the organism, one can learn the normal pathway of this carbon compound in that organism. Thus 14C serves as an isotopic "label" or "tracer".

4. The basis of this technique is that the weight of the nucleus of an atom has little or no effect on the chemical properties of that atom. The chemistry of an element and the atoms of which it is made — whatever their atomic weight — is a function of the atomic number of that element. As long as the atom had 6 protons, it is an atom of carbon irrespective of the number of neutrons. Thus while 6 protons and 8 neutrons produce an isotope of carbon, ^{14}C, 7 protons and 7 neutrons produce a totally-different element, nitrogen-14.

A2. CHEMICAL BONDS

- All interactions between atoms are electrical attractions between charges.
- Ionic and covalent bonds hold atoms together to form molecules
- Weak bonds (hydrogen, van der Waals and other types) hold molecules together



► **FIGURE Relative energies of covalent bonds and noncovalent interactions.** Bond energies are determined as the energy required to break a particular type of linkage. Covalent bonds are one to two powers of 10 stronger than noncovalent interactions. The latter are somewhat greater than the thermal energy of the environment at normal room temperature (25 °C). Many biological processes are coupled to the energy released during hydrolysis of a phosphoanhydride bond in ATP.

1. Ionic bonds: Compounds with ionic bonds split into ions in water. Ions conduct electricity. Gives specialized cells (nerve, muscle) excitable properties. Suppose Na gives one of its electrons to Cl; the Na now has a (+) charge and the Cl will have a (-) charge

- These charged atoms are referred to as ions, and since they have opposite charges they attract each other and form a chemical bond (they form NaCl, common table salt)
- Positively charged ions = cations (i.e., Na^+)
- Negatively charged ions = anions (i.e., Cl^-)

2. Covalent bonds: Each atom donates 1 or more electrons to the bond. The bonding electrons spend most of their time between the 2 atoms, attracting both nuclei and pulling them together. If each atom donates 2 electrons to a bond a double bond is formed

- double bonds are stronger and more rigid than single bonds

A triple bond is formed when each atom donates 3 electrons to the bond. This type of bond holds together the long chains of macromolecules. These molecules do not split apart in water.

3. Hydrogen bonds: Occur when a hydrogen ion is sandwiched between 2 atoms, usually nitrogen and oxygen. It is much weaker (about 25 times) than covalent or ionic bonds and mainly occur between molecular groups with permanent dipoles.

In Water hydrogen bonds makes water molecules stick together. It is responsible for many of the strange properties of water such as cohesion, high boiling point.

Hydrogen bonds cause protein chains to spiral and bend, giving unique shapes. In DNA they hold together the 2 chains to form the double helix. Allow chains to "unzip" for replication and transcription.

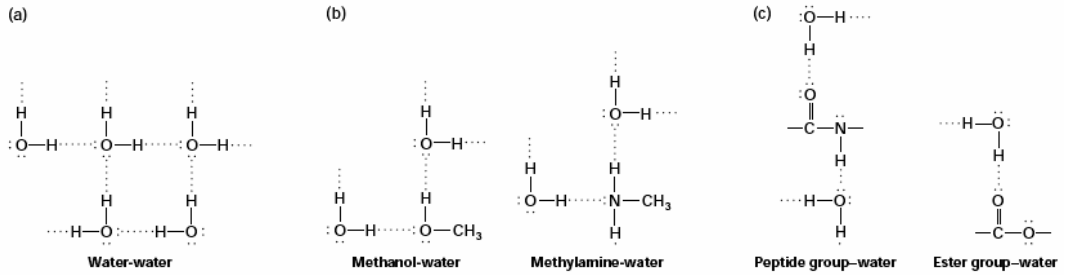


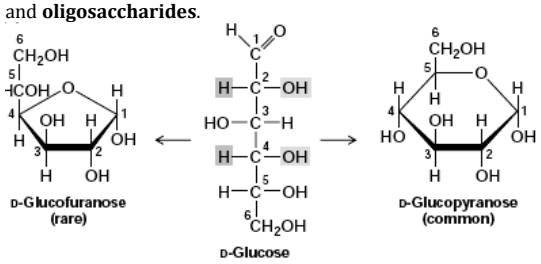
FIGURE Hydrogen bonding of water with itself and with other compounds.

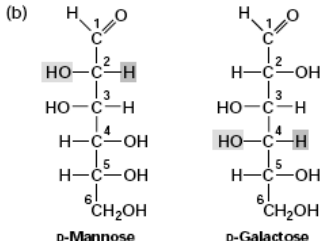
4. Van der Waals & other weak bonds: Weak forces that can bond like atoms together. They are specially important between chains of carbon atoms. They are although weak, numerous bonds between the chains can add up to produce significant cohesion

Such bonds determine physical state of compounds: gas, liquid or solid and occur when one atom induces a temporary dipole in another atom. These are important in holding like molecules together. Often determine the solid, liquid or gas state of a compound. Saturated fats are solid at room temperature because the have more van der Waals attractions than unsaturated fats, which are liquid.

B1. Composition, Structure and function of Carbohydrates

Carbohydrates are mainly compounds of carbon, hydrogen and oxygen. Carbohydrates are so called because in most of them, the proportion of hydrogen and oxygen is the same as in water (H_2O) i.e., 2:1. Their general formula is $C_XH_{2X}O_X$. These are also knows as saccharides (compounds containing sugar). Carbohydrates are produced by green plants during photosynthesis. These constitute about 80% of the dry weight of plants. Carbohydrates are divided into 3 main classes – **monosaccharides**, **derived monosaccharides** and **oligosaccharides**.





1. Monosaccharides

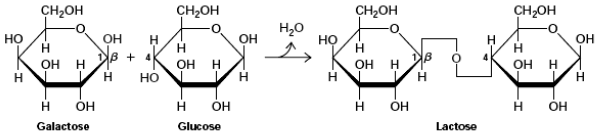
- These have single saccharide units which cannot be hydrolysed further into still smaller carbohydrates; have general formula $C_n(H_2O)_n$. These are composed of 3-7 carbon atoms, and are classified according to the number of C atoms as trioses (3C), tetroses (4C), pentoses (5C), hexoses (6C) and heptoses (7C). Of these, pentoses and hexoses are most common. Monosaccharides are important as energy source and as building blocks for the synthesis of large molecules.
- All monosaccharides are either polyhyroxy aldoses or ketoses. The two simplest monosaccharides are trioses e.g., glyceraldehydes and dilhydroxyacetone.
- Tetroses (e.g. erythrose) are quite rare. Erythrose takes part in the synthesis of lignin and anthocyanin pigments.
- Ribose, ribulose, xylulose and arabinoses are pentoses. Xyluloses and arabinoses polymerise to form xylans and arabans which are cell wall materials.
- Glucose, fructose, mannose, galactose are **hexoses**. These are white, sweet-tasting, crystalline and extremely, soluble in water.
- **Glucose** is the universal sugar. It is also known as **dextrose** or **grape sugar** or **corn sugar**.
- **Fructose** is the most common form of sugar in fruit. It is also known as **levulose**. **It is the sweetest among naturally occurring sugars**.
- Monosaccharides have 'free' aldehyde or ketone group which can reduce Cu^{++} to Cu^{+}. Hence, these are also called reducing sugars.

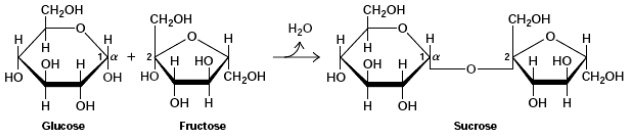
2. Derived monosaccharides

- **Deoxy sugar** – Loss of oxygen atom from ribose yields deoxyribose, a constituent of DNA.
- **Amino sugar** – Monosaccharides having an amino group e.g., glucosamine, galactosamine.
- **Sugar acid** – e.g., Ascorbic acid, glucuronic acid, galacturonic acid.
- **Sugar alcohol** – e.g. glycerol and mannitol (present in brown algae).

3. **Oligosaccharides**: The sugars with limited numbers (2-10) of monosaccharides are called **oligosaccharides**. These include trisaccharides, tetra saccharides, hexasaccharides, heptasaccharides etc.

- **Disaccharides**: These are formed by condensation reactions between two monosaccharides (usually hexoses). The bond formed between two monosaccharides is called a **glycosidic bond**. It normally forms between C-atoms 1 and 4 of neighbouring units (1,4 bond). Once linked, the monosaccharide units are called residues.
  (a) **Maltose** (glucose + glucose), **lactose** (glucose + galactose), sucrose (glucose + fructose) are most common disaccharides. Sucrose is most abundant in plants and is known as **cane sugar** or **table sugar**. It is the sugar we buy from market.
  (b) On **hydrolysis**, disaccharides release their respective constituent monosaccharides (e.g., hydrolysis of sucrose yields one molecule each of glucose and fructose).





- **Trisaccharides**: Sugars composed of 3 monosaccharide units are called **trisaccharides (e.g., raffinose)**. Raffinose is a common saccharide found in plants. Upon hydrolysis, it yields one molecule each of glucose, fructose and galactose.

- **Larger oligosaccharides are attached to the cell membrane and cell recognition is due to their presence. They also take part in** antigen specificity**.**

Polymerisation of a large number of small molecules results in the formation of large molecules of high molecular weights, which may be branched or unbranched. These are

called **macromolecules**. These include some polysaccharides, proteins and nucleic acids.

4. Polysaccharides: These are polymers of monosaccharides and are branced or unbranched linear molecular chains. These are insoluble carbohydrates and are considered to be non-sugars. Starch, glycogen, cellulose, pectin, hemicellulose, inulin and polysaccharides.

5. Polysaccharides are of two types –

- **Homopolysaccharides** – consists of only one type of monosaccharide monomer e.g. starch, glycogen and celluse, fructan, xylan, araban, galactan.
- **Heteropolysaccharide** – consists of more than one type of monosaccharide monomer e.g. chitin, agar, arabanogalactans, arabanoxylan etc.

6. Polysaccharides are of three main types – storage (e.g. starch and glycogen), **structural** (e.g. chitin, cellulose) and **mucopolysaccharides** (e.g. keratin sulphate, chondroitin sulphate, hyaluronic acid, agar, aliginic acid, carragenin & heparin).

7. Storage polysaccharide

(i) **Starch, glycogen** and **inulin** are reserve food materials. Starch turns blue with **Iodine**.

- Starch is a polymer of glucose. It is the major reserve food in plants. Starch has two components – amylase (an unbranched polymer) and amylopectin (a branched polymer).
- **Amylopectin** – consists of 2000 – 200,000 glucose molecules forming straight chain and branches (after 25 glucose units). Branching point has α, 1-6 glycosidic linkage.
- **Amylose** – consists of α, 1-4 glycosidic linkage between α-D glucose molecules. It is straight chain of 200-1000 glucose units. It is helical; each turn consists of 6 glucose units.

Starch molecules accumulate in the form of layers (stratifications) around a shifting organic centre (hilum) to form starch grains. In **eccentric** starch grains, hilum lies on one side. These are found in potatoes. In **concentric** starch grains, hilum is present in the centre. These are found in wheat, maize, and pea.

- **Dumb-bell shaped** starch grains are found in the latex of *Euphorbia*.
- Starch grains with single hilum are called simple (e.g. maize); but those with more than one hilum are called compound (e.g. potato, rice).

(ii) **Glycogen**: Glycogen is the animal equivalent of starch; many fungi also store it. Glycogen turns red-violet with iodine. It consists of 30,000 glucose units joined by α, 1-4 bonds, much more banched than starch. Branching point has α, 1-6 linkage, branching occurs after 10-14 glucose units.

(iii) **Inulin**: It is an unusual polysaccharide and polymer of fructose. It is used as a fructose, particularly in roots and tubers of the family Compositae (e.g. *Dahlia* tubers).

8. **Structural polysaccharide**

a) **Cellulose**: Cellulose is the main structural polysaccharide of plants. It is the most abundant molecule on earth. It is almost confined to plants, though it is found in primitive fungi and lower invertebrates also. It is structural component of all plant cell walls, constituting, on an average, about 20-40% of the wall. Cotton fibres contain the largest proportion (90%) of cellulose among natural materials.

Molecules are unbranched, consisting of 6000 - β – D glucose units joined by β, 1-4 linkages, and 2000 cellulose molecules from **microfibril**. Rayon and cellophane are similar to **cellulose xanthate**. Cellulose acetate is used to prepare tricot, double knit, wrinkle proof and mouth proof clothing, cigarette filters. Cellulose nitrate is used in propellent explosives. Carboxy methyl cellulose is used as emulsifier and smoothing reagent.

b) **Chitin**: Chitin is a polymer of acetyglucosamine (β, 1-4 glycosidic linkage) forming bundles of long parallel chains like cellulose. It occurs in the cell walls of fungi (and forms exoskeleton in some animals especially arthropods.

c) **Pectin and hemicellulose**: Pectin and hemicelluloses are structural polysaccharides. Pectins are made up of **arabinose, galactose** and **galacturonic acid**. Pectic acid is a homopolymer of the methyl ester of D-galacturonic acid.

- **Middle lamella** which binds the cells together is composed of calcium pectate. Due to this substance, water absorption capacity of wall is increased.
- Fruit walls contain high percentage of pectin. During ripening, the pectins break down into sugars resulting in the sweetning and loosening of fruits.
- **Hemicelluloses** are homopolymers of D-xylose linked by β 1-4, Xylans, arabans, galactans are hemicelluloses. These are rarely used as food (e.g. dates – *Phoenix*).

9. Mucopolysaccharides: These are gelatinuous polysaccharides formed from galactose and mannose. Slimy substances of bhindi, agar agar, alginic acid and carrageenin obtained from seaweeds are mucopolysaccharides. Mycopolysaccharides are found in the cell walls of bacteria also.

a) **Keratin sulphate** – consists of acetyl glucosamine, glaactose and sulphuric acid, provides strength and flexibility to skin and cornea.

b) **Chondroitin sulphate** – consists of flycuronic acid and acetyl glucosamine, present in the vitreous humor of eye, synovial fluid and cerebrospinal fluid etc.

c) **Heparin** is a polymer of α, 1-4 glucosamine and glucuronic acid. It is anticoagulant present in human blood. Husk of *Plantago ovata* and mucilage of *Aloe barbedense* are medicinally used.

Agar, alginic acid carragenin are obtained from marine Brown algae(Phaeophyceae).

B2. Lipids

Lipids are esters of fatty acids and polyhydric alcohol. The term 'lipid' was first used by **Bloor** (1943). These are the compounds of C, H, O but the ratio of H and O is more than 2:1 (i.e., the ratio of oxygen is lesses as compared to carbohydrates). These are water-insoluble organic substances which can be extracted from the cells by organic solvents such as ether, chloroform and benzene. Their general formula is $C_nH_{2n}O_2$. Some lipids have P, N and S also.

1. **Simple lipids** – esters of fatty acids with alcohol. Simplest alcohol in fats is glycerol (a trihydric alcohol) e.g., fats, oils and waxes. Triglycerides are common in nature.
2. **Compound lipids** – These lipids contain an additional group alongwith fatty acids and alcohols, e.g., phospholipids, glycolipids and lipoproteins.

3. **Derived lipids** – These are isoprenoid structures e.g., steroids, terpenes, carotenoids.

4. Fatty acids are carboxylic acid with a chain of more than four carbon atoms ending with COOH group. Plants can synthesize all fatty acids. Animals can not synthesize **linoleic, linolenic** and **arachidonic acid**. These are called **essential fatty acids**. Their deficiency causes sterility, kidney failure and stunted growth.

TABLE Fatty Acids That Predominate in Phospholipids

Common Name of Acid (Ionized Form in Parentheses)	Abbreviation	Chemical Formula
SATURATED FATTY ACIDS		
Myristic (myristate)	C14:0	$CH_3(CH_2)_{12}COOH$
Palmitic (palmitate)	C16:0	$CH_3(CH_2)_{14}COOH$
Stearic (stearate)	C18:0	$CH_3(CH_2)_{16}COOH$
UNSATURATED FATTY ACIDS		
Oleic (oleate)	C18:1	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COOH$
Linoleic (linoleate)	C18:2	$CH_3(CH_2)_4CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$
Arachidonic (arachidonate)	C20:4	$CH_3(CH_2)_4(CH{=}CHCH_2)_3CH{=}CH(CH_2)_3COOH$

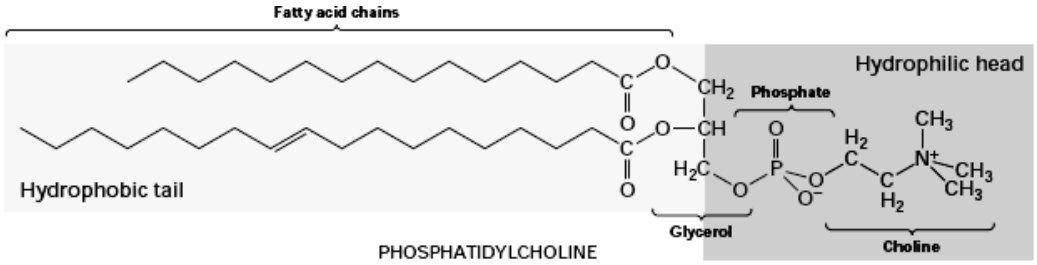
5. Saturated fatty acids have no double bond. Their melting point is high. Palmitic acid, stearic acid are saturated fatty acids. Unsaturated fatty acids are commonly present in vegetable oils, cod/shark oil. Their melting points are low. Oleic acid has one double bond, linoleic acid has two, linolenic acid has three and arachiodonic acid has 4 double bonds. Fatty acids with more than one double bond are called **polyunsaturated fatty acid (PUFA).**

- **Drying oils** are unsaturated fatty acids which can be converted in hard fats on being exposed.
- Edible oils can be converted into hard fats through **hydrogenation**.

6. Waxes are esters of long chain monohydric alcohols like **cetyl, ceryl** or **mericyl.**

- **Lanolin** forms a protective, water insoluble coating on animal fur.
- **Bees wax** secreted from abdominal glands of honey bees has **palmitic acid** and **mericyl alcohol**.
- **Paraffin wax** is a petroleum product.
- **Cutin** is formed by cross esterification and polymerization of hydroxyl fatty acids and other fatty acids without esterification by alcohols other than glycerol. Cuticle has 50-90% cutin.
- **Suberin** is condensation product of **glycerol** and **phellonic acid**. It makes the cell wall impermeable to water.

7. Phospholipids are triglycerides in which one fatty acid is replaced by phosphoric acid which is often linked to additional nitrogenous groups like **choline** (in lecithin), **ethanolamine** (in cepalin), **serine** or **inositol**. These are amphipathic i.e. have both polar and non polar groups. These form cell membranes along with proteins.



8. Sphingolipids have amino alcohol sphingosine. **Sphingomyelins** are present in myelin sheath of nerves. They have additional phosphate attached to choline, are present in nerve membrane. **Cerebrosides** are present in nerve membrane and have **galactose**.

9. Gangliosides – have glucose, galactose, **sialic acid** and **acetyl glucosamine**, present in grey matter, receptors of viral particles, excess causes **Tay-Sachs** disease.

10. Sterols or **steroids** contain 4 fused hydrocarbon rings called **cyclopentane perhydro phenanthrene** and a long side chain e.g. cholesterol, stigmasterol, campesterol, sitosterol, ergasterol.

- Cholesterol helps in absorption of fatty acids, sex hormones, vitamin D and bile salts. Potato is rich in cholesterol. Excess of cholesterol causes **atheroscherosis**.

11. Prostagladins are hormone modulators.

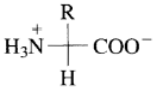
12. Terpenes are lipid like carbohydrates formed of isoprene units e.g., menthol, camphor, carotenoids.

13. Major function of lipids is to act as energy stores.

- Plants usually store oil than fats. Seeds, fruits and choloroplasts are often rich in oils.
- **Glycolipids** are important components of cell membranes, chloroplast membranes.
- Fatty substance in the cell wall (Wax, cutin, suberin) helps reduce transpiration and provide mechanical protection from injury and parasites.
- **Diosgenin** is a steroid obtained from the plant called *Dioscorea*. It is used for manufacturing antifertility pills.

B3. Amino Acids & Proteins

1. Amino acids: Amino acids are building blocks of polypeptides. Polymerization of amino acids forms polypeptides. Amino acids are linked by a peptide bond in a polypeptide. Peptide bond are synthesized during translation of messenger RNA



2. Primary structure of a protein is the sequence of amino acids. Both peptides and polypeptides can be functional

3. Amino acids may be functional. Various amino acids functions as neurotransmitters

- glutamate and aspartate (excitory)
- glycine, taurine, and γ-aminobutyric acid (GABA) (inhibitory)

Structure $^{-}OOC\text{-}CH_2\text{-}CH_2\text{-}NH_3^+$ GABA

Structure $^{+}H_3N\text{-}CH_2\text{-}CH_2\text{-}SO_3^-$ TAURINE

4. Amino Acids can act as precursors to other molecules

- **metabolic intermediates:** citrulline and ornithine in urea cycle can be metabolized to form glucose or acetyl CoA
- **neurotransmitters**- serotonin, dopamine, epinephrine, etc.
- tyrosine (thyroid hormone)
- porphyrins
- creatine (energy storage)
- histamine (mediator of immune response)
- nucleotide synthesis - S-adenosylmethionine

5. Structure of amino acids: All proteins are composed of 20 standard amino acids, which are specified by the genetic code. The standard amino acids are called α-amino acids because they have a primary amino group and a carboxyl group bound to the same carbon atom (the α carbon). Only proline has a secondary amino group attached to the α carbon, but it is still commonly referred to as an α-amino acid. The generic structure of an amino acid at pH 7 is shown below.

At pH 7, the amino acid is a zwitterion, or dipolar ion. A unique side chain, or R group, characterizes each amino acid.

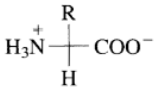
There are 20 standard α-amino acids (although proline is regarded as α-imino acid). An Amino acid has a alpha-carbon, alpha-amino group, alpha-carboxyl group and a side chain (R group). These "standard" amino acids are encoded by messenger RNA. Amino acids are abbreviated by a 3-letter and 1-letter code

Abbreviations for the 20 "standard" amino acids

alanine	Ala	A	leucine	Leu	L
arganine	Arg	R	lysine	Lys	K
asparagines	Asn	N	methioneine	Met	M
aspartic acid	Asp	D	phenylalanine	Phe	F
cysteine	Cys	C	proline	Pro	P
glycine	Gly	G	serine	Ser	S
glutamine	Gln	Q	threonine	Thr	T
glutamic acid	Glu	E	tryptophan	Trp	W
histidine	His	H	tyrosine	Tyr	Y
isoleucine	Ile	I	valine	Val	V

Chemical Properties of Amino Acids

6. **Charge**: Amino acids are dipolar ions (zwitterions) at neutral pH. Zwitterion is a dipolar molecule with positive and negative charges spatially separated. Ionic states of amino acids depend on pH. Amino acids have two or three dissociable protons. pKa of the dissociable proton and the pH determine its degree of dissociation.



7. Amino acids are polymerized by condensation reactions to form a chain called a polypeptide. Each polypeptide is polarized: One end has a free amino group and the other end has a free carboxyl group, referred to as the N-terminus and the C-terminus, respectively.

8. The R groups of the standard 20 amino acids are classified into three categories based on their polarities and charge at pH 7: the nonpolar amino acids, the polar uncharged amino acids, and the charged amino acids.

9. Among the nonpolar amino acids, glycine (shorthand Gly or G) has a hydrogen atom as its R group. Alanine (Ala; A), valine (Val; V), leucine (Leu; L), isoleucine (Ile; I), and methionine (Met; M) have aliphatic chains as R groups (Met has a sulfur rather than a methylene group). Tryptophan (Trp; W) and phenylalanine (Phe; F) contain bulky indole and phenyl groups, respectively. These tend to orient inside of a protein.

10. The polar uncharged amino acids include asparagine (Asn; N), glutamine (Gln; Q), serine (Ser; S), threonine (Thr; T), tyrosine (Tyr; Y), and cysteine (Cys; C). Amide functional groups occur in Gln and Asn. Alcoholic functional groups occur in Ser and Thr. Tyrosine and cysteine are characterized by a phenolic group and a thiol group, respectively.

11. Hydropathicity for an amino acid is a index of solubility characteristics in H_2O. It combines hydrophobic and hydrophilic tendencies, It can be used to predict protein structure

12. Among the charged amino acids, aspartate (Asp; D) and glutamate (Glu; E) contain carboxylic groups in their R groups. Lysine (Lys; K), arginine (Arg; R), and histidine (His; H) contain a butylammonium group, a guanidino group, and an imidazole group, respectively. Both polar uncharged and charged aminacids tend to orient to the outside of proteins

13. The p*K* values of ionizable groups depend on the electrostatic influences of nearby groups. Inside of proteins, the p*K* values of ionizable R groups may shift by several pH units from their values in the free amino acids.

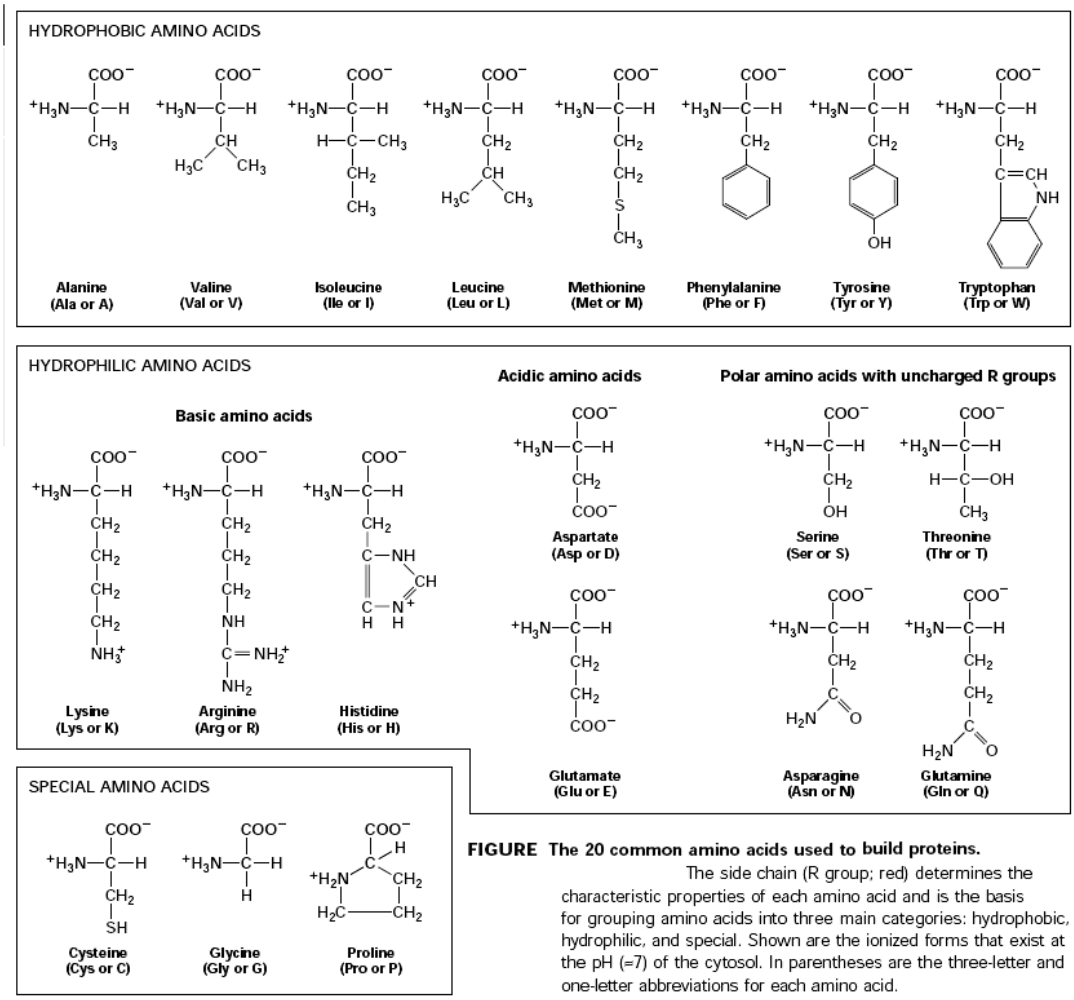


FIGURE The 20 common amino acids used to build proteins. The side chain (R group; red) determines the characteristic properties of each amino acid and is the basis for grouping amino acids into three main categories: hydrophobic, hydrophilic, and special. Shown are the ionized forms that exist at the pH (≈7) of the cytosol. In parentheses are the three-letter and one-letter abbreviations for each amino acid.

Stereochemistry

13. Except for glycine, the standard amino acids have asymmetric structures and rotate the plane of polarized light; thus, they are optically active. These molecules cannot be superimposed on their mirror images. Such nonsuperimposable pairs of molecules are called enantiomers. The asymmetric atom of an optically active molecule is called the chiral center and the molecule is said to have the property of chirality.

14. Fischer projections are used to represent the absolute configuration of substituents around a chiral center. In the Fischer convention, horizontal lines extend above the plane of the paper, while vertical lines extend below the surface of the paper. The α-amino acid shown above is a Fischer projection of an L-amino acid. The specific arrangement of substituents around the chiral carbon is related to that of L-glyceraldehyde.

15. A molecule with *n* chiral centers has 2*n* different possible stereoisomers. Molecules with two or more chiral centers are better described by the *RS* system, in which each substituent bound to the chiral center is prioritized according to its atomic number. Hence, the exact molecular arrangement of a molecule can be unambiguously described.

16. Biochemical reactions almost invariably produce pure stereoisomers, in large part due to the precise arrangement of chiral groups inside of enzymes, which restricts the geometry of the reactants.

17. Nonstandard amino acids in polypeptides arise from posttranslational modifications of the R groups of amino acid residues of the polypeptide. These modifications have critical roles in the structure and function of proteins. Many unpolymerized nonstandard amino acids are synthesized by chemical modifications of one of the standard amino acids. Cells use many of these amino acids as signaling molecules, particularly in the central nervous system.
Example 1: **Hydroxyproline and hydroxylysine:** Proline and lysine are hydroxylated enzymatically after translation. They are important in collagen structure (Structures of phosphoserine, 4-OH-proline, 5-OH-lysine)
Example 2: **Phospho-aminoacids:** Tyr, Ser, Thr hydroxyl groups can be phosphorylated. They are important in activation and inhibition of enzymatic or signalling activity

Among the nonstandard amino acids are also the D isomers of the standard amino acids; many of these occur in bacterial cell walls and bacterially produced antibiotics. The tripeptide glutathione is a cellular reducing agent.

18. Titration curve of an amino acid is calculated using the Henderson Hesselbech equation Isoelectric point (pI) is the pH at which the molecule has a net charge = 0 (average of the two appropriate pKa values)

Key Equation: $pI = \frac{1}{2}(pKi + pKj)$

19. Aromatic amino acids (Trp, Tyr, Phe) absorb UV light between 260-280 nm with maxima at 280 nm.

20. Cysteine can form disulfide bonds. Disulfide bridges formed between cysteines are important in protein structure. Cysteine is the reduced form (sulfhydryl) whereas cystine is the oxidized form (disulfide).

B 4 Protein Structure and Protein Folding

1. Peptide bond – Chemically it is an amide bond between alpha-amino and alpha-carboxyl groups of 2 amino acids. Peptide bond is polar and planar electron resonance structure has partial (40%) double bond character amide group is planar, usually trans. Synthesis of peptide bond involves condensation and water is produced. Energy is required for this process which is obtained by ATP hydrolysis.

2. Peptide bond is hydrolysable:

- Acid hydrolysis with 6N Hydrochloric acid heated at 110^0C for 24 hr in a vacuum generates free amino acids breaking all peptide bonds.
- Base hydrolysis with 4N Sodium hydroxide heated at 100^0C for 4 hr also generates free amino acids.
- Cyanogen bromide cleaves at the COOH-terminal side of Met.
- Enzymatic hydrolysis of peptide bonds by proteases like trypsin cleave after –COOH terminus Lys or Arg; Chymotrypsin after carboxy terminus of aromatic amino acids yield peptides of various length

3. Polypeptides are polyampholytes: Ampholyte has both acidic and basic pKa values. **Isoelectric point** is the pH at which the net charge is zero

For example: H_3N^+-Ala-Lys-Ala-Ala-COO^-

pKa of the Alpha-carboxyl group = 3.6

pKa of the Alpha-amino group = 8.0

pKa of the delta-amino of the Lysine = 10.6

at pH = 1 the net charge is +2

at pH = 6 the net charge is +1

at pH = 14 the net charge is -1

The isoelectric point pI = (pKa2 + pKa3)/2 = (8 +10.6)/2 = 9.3

4. **Nomenclature of Peptides on basis of size**

- dipeptide(2 amino acids & 1 peptide bond), tripeptide(3 amino acids & 2 peptide bonds)
- oligopeptide - several amino acids (up to 20)
- polypeptides (more than 20 amino acids). All proteins are polypeptides

5. Physical Forces Governing Protein Conformation: Physical forces govern 3-D structure of proteins (Pauling and Corey)

- bond lengths and angles should be distorted as little as possible
- structures must follow Van der Waal's rules for atomic radii
- peptide bond is planar and trans
- noncovalent bonding stabilizes structure
- conformation can change without breaking bonds (flexibility)

6. **Types of non-covalent forces important to protein conformation**

- Hydrophobic forces: Hydrophobic residues orient to inside while hydrophilic residues orient out
- Van der Waal's potential: includes electron shell repulsion, dispersion forces, and electrostatic interactions
- Salt bridges, electrostatic forces and Hydrogen bonds

7. Angles of rotation of the polypeptide chain determine structure

- angles of rotation around alpha-carbon are (ψ psi and φ phi). ψ (psi) is the angle of the alpha-carbon bond to the carbonyl-carbon. φ (phi) is the angle of the alpha-carbon bond to the amide-nitrogen.
- primary sequence and angles of rotation for each alpha-carbon completely define protein conformation
- only a small number of psi and phí angles are allowed. Statistical analysis of all proteins yields groups of prefered angles. Areas of repeating (psi) and (phi) angles are secondary structures

8. Primary Structure of protein is amino acid sequence of a polypeptide. Primary structure determines 3-dimensional structure. Primary structure is always represented NH_2-terminus to COOH-terminus

9. Secondary structure is formed by regular local conformation of linear segments of the polypeptide chain. Secondary structure are stabilized by hydrogen bonds between amide and carbonyl groups. Several types of secondary structure

- **alpha-helix:** It is right handed helix with 3.6 amino acids per turn, rise per helix 5.4 A°, rise per amino acid 1.5 A°. Here carbonyl oxygen hydrogen of one amino acid is bonded to 4th amide hydrogen (n→n+4). Amino acids R-groups orient out. Mostly proline breaks the helix.
- **beta-pleated sheet:** Here polypeptide chains side by side. Polypeptide chains can be parallel or anti-parallel. Carbonyl oxygen hydrogen bonded to amide hydrogen. Beta-strand is a single pass of the polypeptide
- **reverse turn, beta-bend** allows a sharp turn in polypeptide chain. Here carbonyl oxygen hydrogen bonded to 3rd amide hydrogen (n→n+3). Glycine is required.

10. Fibrous proteins demonstrate secondary structure.

- **Fibroin:** silk is fibroin which has anti parallel-beta-pleated sheet
- **α-Keratin and tropomyosin:** alpha-keratins occurs in wool and hair and epidermal layer. Tropomyosin is a thin filament in muscle. Both have α-helix. Alpha-helix allows elasticity. α-keratin converts to β-pleated sheet with heat or stretching. Disulfides are important to maintenance of keratin secondary structure. α-alpha-keratins are β-pleated sheets in feathers and claws of birds.
- **Collagen:** It is structural protein found in skin, bones. It has triple helix (not α-helix). It has repeating sequence (Gly-X-Pro) X or (Gly-X-HyPro)X. Glycine required for triple helix to form, contains many modified amino acids. Hydroxyproline stabilizes the structure, vitamin C required for hydroxylation thus in Vitamin C deficiency cause **scurvy**.

11. Tertiary structures are overall folded conformation of the polypeptide. Physical forces affect tertiary structure. The main driving force is **hydrophobic force**, as a result,

hydrophobic residues orient to inside and hydrophilic orient out. To some extent salt bridges, electrostatic forces, Van der Waals radii, Hydrogen bonds and Disulfide bridges also play important role in maintaining tertiary structure. Myoglobin has a tertiary structure.

12**. Quaternary structure** are formed by aggregation of 2 or more subunits. These may be hetero- or homo- polymers. Same forces drive tertiary and quaternary structure. Haemoglobin has quaternary structure.

13. **Protein folding:** Folding is driven by **hydrophobic forces.** Folding occurs step-wise with several intermediates. Unfolded→secondary structure→domains→molten globule→native tertiary structure
A collapsed structure (molten globule) occurs very quickly. Steps between molten globule and native tertiary structure usually occur slowly. The intermediates are isolatable and there may be multiple pathways for folding.

14. Proteins can self assemble but in vivo folding is facilitated by proteins. **Chaperones** are binding proteins which assist folding. Chaperones cause misfolded protein to unfold rather than aggregate. Many chaperones require ATP hydrolysis for activity. More than one chaperone may act simultaneously and sequentially in the folding of a single protein. Chaperones are specific for specific protein synthesis pathways (cytosolic vs. mito. vs. endoplasmic reticulum)

15. Enzymes catalyze kinetically slow steps in folding

- **cis-trans prolyl isomerase:** Both cis and trans peptide bonds to proline naturally occur. The isomerization of peptidyl-proline bonds may be slow so this enzyme catalyzes reaction at enhanced rate.
- **protein disulfide isomeras**e: catalyzes disulfide bond formation and isomerization

16**. Denaturation is unfolding:** It requires some input to overcome hydrophobic forces. The heat or other denaturant (urea or guanidinium) can unfold the proteins. For reducing disulfide bridges to sulfhydryls certain reductant (Meracaptoethanol) are required.

B5. Nucleic Acid

1. Nucleic acids (DNA/RNA) are polymer of nucleotides. As a class, the nucleotides may be considered one of the most important metabolites of the cell. Nucleotides are found primarily as the monomeric units comprising the major nucleic acids of the cell, RNA and DNA. However, they also are required for numerous other important functions within the cell. These functions include:

a. serving as energy stores for future use in phosphate transfer reactions. These reactions are predominantly carried out by ATP.
b. forming a portion of several important coenzymes such as NAD+, NADP+, FAD and coenzyme A.
c. serving as mediators of numerous important cellular processes such as second messengers in signal transduction events. The predominant second messenger is cyclic-AMP (cAMP), a cyclic derivative of AMP formed from ATP.
d. controlling numerous enzymatic reactions through allosteric effects on enzyme activity.
e. serving as activated intermediates in numerous biosynthetic reactions. These activated intermediates include S-adenosylmethionine (SAM) involved in methyl transfer reactions as well as the many sugar coupled nucleotides involved in glycogen and glycoprotein synthesis.

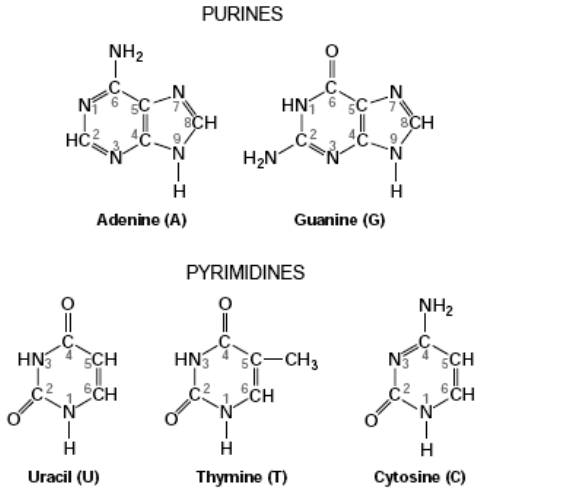


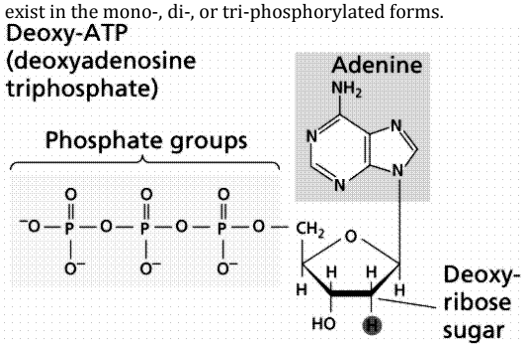
FIG Chemical structures of the principal bases in nucleic acids.

2. The nucleotides found in cells are derivatives of the heterocyclic highly basic, compounds, purine and pyrimidine.

It is the chemical basicity of the nucleotides that has given them the common term "**bases**" as they are associated with nucleotides present in DNA and RNA. There are five major bases found in cells. The derivatives of purine are called adenine and guanine, and the derivatives of pyrimidine are called thymine, cytosine and uracil. The common abbreviations used for these five bases are, A, G, T, C and U.

3. The purine and pyrimidine bases in cells are linked to carbohydrate and in this form are termed, **nucleosides**. The nucleosides are coupled to D-ribose or 2'-deoxy-D-ribose through a b-N-glycosidic bond between the anomeric carbon of the ribose and the N9 of a purine or N1 of a pyrimidine.

4. Nucleosides are found in the cell primarily in their phosphorylated form. These are termed nucleotides. The most common site of phosphorylation of nucleotides found in cells is the hydroxyl group attached to the 5'-carbon of the ribose The carbon atoms of the ribose present in nucleotides are designated with a prime (') mark to distinguish them from the backbone numbering in the bases. Nucleotides can exist in the mono-, di-, or tri-phosphorylated forms.



5. Nucleotides are given distinct abbreviations to allow easy identification of their structure and state of phosphorylation. The monophosphorylated form of adenosine (adenosine-5'-monophosphate) is written as, AMP. The di- and tri-phosphorylated forms are written as, ADP and ATP, respectively. The use of these abbreviations assumes that the nucleotide is in the 5'-phosphorylated form. The di- and tri-phosphates of nucleotides are linked by acid anhydride

bonds. Acid anhydride bonds have a high ΔG^0 for hydrolysis imparting upon them a high potential to transfer the phosphates to other molecules. It is this property of the nucleotides that results in their involvement in group transfer reactions in the cell.

6. The nucleotides found in DNA are unique from those of RNA in that the ribose exists in the 2'-deoxy form and the abbreviations of the nucleotides contain a d designation. The monophosphorylated form of adenosine found in DNA (deoxyadenosine-5'-monophosphate) is written as dAMP.

7. The nucleotide uridine is never found in DNA and thymine is almost exclusively found in DNA. Thymine is found in tRNAs but not rRNAs nor mRNAs. There are several less common bases found in DNA and RNA. The primary modified base in DNA is 5-methylcytosine. A variety of modified bases appear in the tRNAs. Many modified nucleotides are encountered outside of the context of DNA and RNA that serve important biological functions.

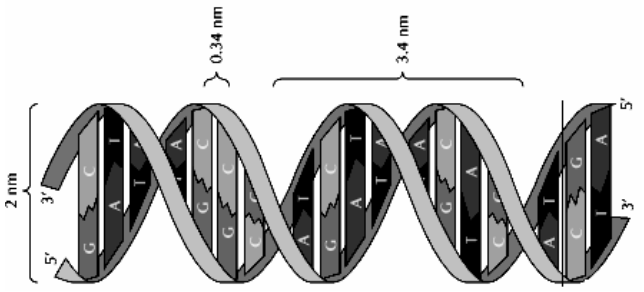
8. **Adenosine Derivatives: 3'-5'-cyclic adenosine monophosphate (cAMP)** is a very powerful second messenger involved in passing signal transduction events from the cell surface to internal proteins, e.g. cAMP-dependent protein kinase (PKA). Cyclic-AMP is also involved in the regulation of ion channels by direct interaction with the channel proteins, e.g. in the activation of odorant receptors by odorant molecules. Formation of cAMP occurs in response to activation of receptor coupled adenylate cyclase. These receptors can be of any type, e.g. hormone receptors or odorant receptors. **S-adenosylmethionine(SAM)** is a form of activated methionine which serves as a methyl donor in methylation reactions and as a source of propylamine in the synthesis of polyamines.

9. Guanosine Derivatives: cyclic form of GMP (cGMP) also is found in cells involved as a second messenger molecule. In many cases its role is to antagonize the effects of cAMP. Formation of cGMP occurs by guanylate cyclase that is coupled to the receptor. The most important cGMP coupled signal transduction cascade is that photoreception.

10. Synthetic Nucleotide Analogs: Large family of analogues is used as anti-tumor agents, for instance, because they interfere with the synthesis of DNA and thereby preferentially kill rapidly dividing cells such as tumor cells. Some of the nucleotide analogues commonly used in chemotherapy are 6-mercaptopurine, **5-fluorouracil, 5-iodo-2'-deoxyuridine and 6-thioguanine**. Each of these compounds disrupts the normal replication process by interfering with the formation of correct Watson-Crick base-pairing. Nucleotide analogs also have been targeted for use as antiviral agents. Several analogs are used to interfere with the replication of HIV, such as **AZT (azidothymidine) and ddI (dideoxyinosine).**

Several purine analogs are used to treat gout. The most common is **allopurinol,** which resembles hypoxanthine. Allopurinol inhibits the activity of xanthine oxidase, an enzyme involved in de novo purine biosynthesis. Additionally, several nucleotide analogues are used after organ transplantation in order to suppress the immune system and reduce the likelihood of transplant rejection by the host.

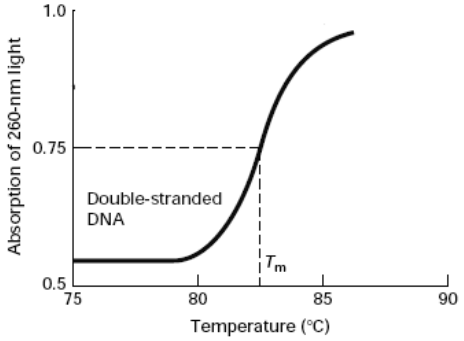
11. Polynucleotides: Polynucleotides are formed by the condensation of two or more nucleotides. The condensation most commonly occurs between the alcohol of a 5'-phosphate of one nucleotide and the 3'-hydroxyl of a second, with the elimination of H_2O, forming a phosphodiester bond. The formation of phosphodiester bonds in DNA and RNA exhibits directionality. The primary structure of DNA and RNA (the linear arrangement of the nucleotides) proceeds in the 5' → 3' direction. The common representation of the primary structure of DNA or RNA molecules is to write the nucleotide sequences from left to right synonymous with the 5' → 3' direction as shown: **5'-pGpApTpC-3'**



12. STRUCTURE OF DNA: Utilizing X-ray diffraction data, obtained from crystals of DNA, James Watson and Francis Crick proposed a model for the structure of B-DNA. This model predicted that DNA would exist as a helix of two complementary antiparallel strands, wound around each other in a rightward direction and stabilized by H-bonding between bases in adjacent strands. In the Watson-Crick model, the bases are in the interior of the helix aligned at a nearly 90 degree angle relative to the axis of the helix. Purine bases form hydrogen bonds with pyrimidines, in the crucial phenomenon of base pairing. Experimental determination has shown that, in any given molecule of DNA, the concentration of adenine (A) is equal to thymine (T) and the concentration of cytidine (C) is equal to guanine (G). This means that A will only base-pair with T, and C with G. According to this pattern, known as Watson-Crick base-pairing, the base-pairs composed of G and C contain three H-bonds, whereas those of A and T contain two H-bonds. This makes G-C base-pairs more stable than A-T base-pairs.

The antiparallel nature of the helix stems from the orientation of the individual strands. From any fixed position in the helix, one strand is oriented in the 5' →3' direction and the other in the 3' → 5' direction. On its exterior surface, the double helix of DNA contains two deep grooves between the ribose-phosphate chains. These two grooves are of unequal size and termed the major and minor grooves. The difference in their size is due to the asymmetry of the deoxyribose rings and the structurally distinct nature of the upper surface of a base-pair relative to the bottom surface.

13. Thermal Properties of DNA: If a solution of DNA is subjected to high temperature, the H-bonds between bases become unstable and the strands of the helix separate in a process of thermal denaturation. Regions of the duplex that have predominantly A-T base-pairs will be less thermally stable than those rich in G-C base-pairs. In the process of thermal denaturation, a point is reached at which 50% of the



DNA molecule exists as single strands. This point is the ***melting temperature (Tm),*** and is characteristic of the base composition of that DNA molecule. The TM depends upon several factors in addition to the base composition. These include the chemical nature of the solvent and the identities and concentrations of ions in the solution.

14 DNA: The Genetic Material --The physical carrier of inheritance

- **Friedrich Meischer** in 1869 isolated DNA from fish sperm and the pus of open wounds. Since it came from nuclei, Meischer named this new chemical, nuclein. Subsequently the name was changed to nucleic acid and lastly to deoxyribonucleic acid (DNA). Robert Feulgen, in 1914, discovered that fuchsin dye stained DNA. DNA was then found in the nucleus of all eukaryotic cells.

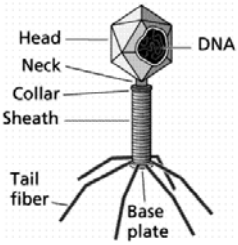
- During the 1920s, biochemist **P.A. Levene** analyzed the components of the DNA molecule. He found it contained four nitrogenous bases: cytosine, thymine, adenine, and guanine; deoxyribose sugar; and a phosphate group. He concluded that the basic unit (nucleotide) was composed of a base attached to a sugar and that the phosphate also attached to the sugar. He (unfortunately) also erroneously concluded that the proportions of bases were equal and that there was a tetranucleotide that was the repeating structure of the molecule. The nucleotide, however, remains as the fundemantal unit (monomer) of the nucleic acid polymer. There are four nucleotides: those with cytosine (C), those with guanine (G), those with adenine (A), and those with thymine (T).

- During the early 1900s, the study of genetics began in earnest: the link between Mendel's work and that of cell biologists resulted in the chromosomal theory of inheritance; **Garrod** proposed the link between genes and "inborn errors of metabolism"; and the question was formed: what is a gene? The answer came from the study of a deadly infectious disease: pneumonia.

- During the 1920s **Frederick Griffith** studied the difference between a disease-causing strain of the pneumonia causing bacteria (Streptococcus peumoniae) and a strain that did not cause pneumonia. The pneumonia-causing strain (the S strain) was surrounded by a capsule. The other strain (the R strain) did not have a capsule and also did not cause pneumonia. Frederick Griffith (1928) was able to induce a nonpathogenic strain of the bacterium Streptococcus pneumoniae to become pathogenic. Griffith referred to a transforming factor that caused the non-pathogenic bacteria to become pathogenic. Griffith injected the different strains of bacteria into mice. The S strain killed the mice; the R strain did not. He further noted that if heat killed S strain was injected into a mouse, it did not cause pneumonia. When he combined heat-killed S with Live R and injected the mixture into a mouse (remember neither alone will kill the mouse) that the mouse developed pneumonia and died. Bacteria recovered from the mouse had a capsule and killed other mice when injected into them.

- **Hypotheses:**
  1. The dead S strain had been reanimated/resurrected.
  2. The Live R had been transformed into Live S by some "transforming factor".
  Further experiments led Griffith to conclude that number 2 was correct.

- **In 1944, Oswald Avery, Colin MacLeod, and Maclyn McCarty** revisited Griffith's experiment and concluded the transforming factor was DNA. Their evidence was strong but not totally conclusive. The then-current favorite for the hereditary material was protein; DNA was not considered by many scientists to be a strong candidate.

- 

The breakthrough in the quest to determine the hereditary material came from the work of **Max Delbruck and Salvador Luria** in the 1940s. Bacteriophage are a type of virus that attacks bacteria, the viruses that Delbruck and Luria worked with were those attacking Escherichia coli, a bacterium found in human intestines. Bacteriophages consist of protein coats covering DNA. Bacteriophages infect a cell by injecting DNA into the host cell. This viral DNA then "disappears" while taking over the bacterial machinery and beginning to make new virus instead of new bacteria. After 25 minutes the host cell bursts, releasing hundreds of new bacteriophage. Phages have DNA and protein, making them ideal to resolve the nature of the hereditary material.

- In 1952, **Alfred D. Hershey and Martha Chase** conducted a series of experiments to determine whether protein or DNA was the hereditary material. By labeling the DNA and protein with different (and mutually exclusive) radioisotopes, they would be able to determine which chemical (DNA or protein) was getting into the bacteria. Such material must be the hereditary material (Griffith's transforming agent). Since DNA contains Phosphorous (P) but no Sulfur (S), they tagged the DNA with radioactive Phosphorous-32. Conversely, protein lacks P but does have S, thus it could be tagged with radioactive Sulfur-35. Hershey and Chase found that the radioactive S remained outside the cell while the radioactive P was found inside the cell, indicating that DNA was the physical carrier of heredity.

15. Polymorphism in DNA:

- **A form** – having 11 base pairs (instead of 10 base pairs per turn), the base pairs are not perpendicular to the axis, but are tilted.
- **C form** – like B form, but having 9 base pairs per turn.
- **D form** – like B form, but have 8 base pairs per turn.
- **Z-DNA**-DNA with left handed coiling is called **Z-DNA**.

16. DNA supercoiling: A supercoil is when the double-helix twists around itself. Supercoils can be positive or negative but natural DNAs exist in the **negative supercoiled form**. DNA can be supercoiled if it is circular or if is linear and has fixed ends. Supercoiled DNA is more compact than relaxed DNA. Negatively supercoiled DNA molecules are easier to unwind than relaxed molecules. DNA unwinding is required for replication and transcription. **Topoisomerases** are enzymes that catalyze changes in DNA supercoiling.

a. ***Type I topoisomerases*** function by breaking a phosphodiester bond of one strand, passing the other strand through the break and resealing the break. They can only remove supercoil and increase linking number by 1.

b. ***Type II topoisomerases*** function by breaking both strands and passing a double strand region through the break before resealing the break, for this process they require ATP. They dectrease linking number by 2.

Topoisomerases are targets of numerous chemotherapeutic drugs: **adriamycin**, VP16 (**tenoposide**), VM26 (**etoposide**), **camptothecin.**

17. **Hyperchromic effect:** On denaturation of dsDNA into SS DNA UV absorption at 260 nm increases.

18. Packaging of DNA in eukaryotes: DNA is packaged in the nucleus as a nucleoprotein complex called chromatin. Levels of chromatin packaging:

a. Nucleosomes: Nucleosome core particles consist of ~140 bp of DNA wrapped around a protein octamer consisting of 2 subunits each of histones H2A, H2B, H3, and H4. Linker DNA (~60 bp) is the DNA between two core particles, histone H1 binds to the linker DNA and the core particle.

b. **30 nm fiber(Solenoid):** nucleosomes are wound into a solenoid-like structure. It requires histone H1 binding to every linker DNA

c. chromatin is the template for replication and transcription and the substrate for DNA repair and recombination

B6. NUTRITION AND VITAMINS:

Nutrition is the process of acquiring energy and materials for cell metabolism including the Maintenance and repair of cells, and growth.

HETEROTROPHIC NUTRITION:

1. Hetrotrophic organisms or hetrotrophs are organisms that feed on an organic source of carbon. The survival of hetrotrophs is depend either directly or indirectly on the activities of autotrophs. A few bacteria are able to use light energy to synthesise their organic requirements from other raw materials. These are called **photoautotrophs.**

2. The way in which hetrotrophs obtain their food varies considerably. However the way in which it is processed into a usable form within the body is very similar in most of them. It involves following processes:

- **Digestion:** This is the process which reduce large complex food molecules into simpler soluble ones.
- **Absorption:** In this process the soluble molecules are taken away from the region of digestion into the tissues of the organisms.
- **Assimilation:** It use the absorbed nutrients for a particular purpose.

3. Forms of Hetrotrophic Nutrition

- **Holozoic Nutrition**
- **Saprotrophic Nutrition**

4. **Holozoic Nutrition:** The term **holozoic** is applied mainly to free living animals which have a specialised digestive tract, the alimentary canal. Most animals are holozoic. The characteristic processes involved in holozoic nutrition are as follows:

- **Ingestion** is the taking in of food.
- **Digestion** is the breakdown of large organic molecules into smaller, simpler soluble molecules. Often two types of digestion occur. ***Mechanical digestion*** involves mechanical breakdown of the food, for example by teeth. ***Chemical digestion*** involves the activity of enzymes. The type of chemical process these enzyme catalyse during digestion is hydrolysis. Digestion may be extracellular or intracellular.
- **Absorption:** It is the uptake of the soluble molecules from the digestive region, across a membrane and into the body tissue proper. The food may pass directly into cells or first pass into the blood stream to be transported to other regions of the body.
- **Assimilation** is using the absorbed molecules to provide either energy or materials to be incorporated into the body.
- **Egestion** is the elimination from the body of undigested waste food materials. Animals which feed on plants are called as herbivores, those that feed on animals are called as carnivores, and those that eat a mixed diet of animal and vegetable matter are termed **omnivores.** Some animals take food in the form of relatively small particles, for ex earthworms and filter feeders like mussels. Some ingest food into liquid form ex., aphids, butterflies and mosquitoes

5. Saprotrophic Nutrition (sapros = Rotten, Trophos = Decay): Organisms which feed on dead or decaying organic matter are called saprotrophs other terms which mean the same thing is saprophytes and saprobiant. Many fungi and bacteria are saprotrophs eg. *mucor*, *rhizophus*, yeasts. Saprotrophs secrete enzymes onto their food where it is digested .The soluble end products of this extracellular digestion are than absorbed and assimilated by the saprotrophs. Saprotrophs feed on the dead organic remains of plants and animals and contribute to the removal of such remains by decomposing it. Many of simple substances are not used by the saprotrophs by themselves but are absorbed by plants. In this way the saprotroph provide important links in nutrient cycles by making possible the return of vital chemical elements from the dead bodies organisms to living ones.

6. Saprotophic nutrition of Mucor and Rhizophus: *Mucor* and *Rhizophus* are common fungi known as pin moulds. They are often found growing on bread, although they can also live in soil. Their hyphae penetrate the food on which they grow and secrete hydrolyzing enzyme from their tips. Enzymes carbohydrase and protease carry out the extracellular digestion of starch to glucose and protein to amino acids respectively. The thin much branched nature of the mycelium of *mucor* and *rhizophus* ensures that there is a large surface area for absorption. Glucose is used during respiration to provide energy for the organisms metabolic activities whilst glucose and amino acids are used for the growth and repair. Excess glucose is converted to glycogen and fat and excess amino acid to protein granules for storage in cytoplasm.

7. Symbiosis: The term symbosis means literally living together. It was introduced by a german scientist de Bary in 1879,who described it as living together of dissimilarly named organisms in other words it is in association between two or more organisms of different species The symbiosis has been restricted by many biologists to meaning a close relationship between two or more organisms of different spp. In which all partners benefit. Since 1970 symbiosis is more important as a topic in biology. For example we now know that the great majority of plants obtain minerals with assistance of fungi and that much nitrogen fixation is carried out by symbiotic bacteria. Symbiosis is the living together in close association of two or more organisms of different species many associations involve three or more partners. Nutrition is commonly involved .

8. Three Common Types of Symbiotic Relationships are

- **Mutualism-** In which both Partners Benefit
- **Parasitism-**In which one partner benefits and causes harm to other
- **Commensalism**-In which one partner benefits but the other recives no harm or benefit.

NUTRITION IN HUMANS

9. We need food for energy not only because it tastes good and that gives us a great feeling when we have it, but also because it gives us the much needed energy required to carry out our routine activities.

- Human beings require food to grow, reproduce, and maintain good health. Without food, our bodies could not stay warm, build or repair tissue, or maintain a heartbeat.
- Eating the right foods can help us avoid certain diseases or recover faster when illness occurs.
- These and other important functions are fueled by chemical substances in our food called nutrients. Nutrients are classified as carbohydrates, proteins, fats, vitamins, minerals, and water.

10. When we eat a meal, nutrients are released from food through digestion. Digestion begins in the mouth by the action of chewing and the chemical activity of saliva, a watery fluid that contains enzymes, certain proteins that help break down food. Further digestion occurs as food travels through the stomach and the small intestine, where digestive enzymes and acids liquefy food and muscle contractions push it along the digestive tract. Nutrients are absorbed from the inside of the small intestine into the bloodstream and carried to the sites in the body where they are needed. At these sites, several chemical reactions occur that ensure the growth and function of body tissues. The parts of foods that are not absorbed continue to move down the intestinal tract and are eliminated from the body as feces.

11. What are the nutrients that provide energy to our body and how it is measured?

- Carbohydrates, proteins, and fats provide the body with the energy it needs to maintain its many functions.
- Scientists measure this energy in kilocalories, the amount of energy needed to raise 1 kilogram of water 1 degree Celsius. In nutrition discussions, scientists use the term calorie instead of kilocalorie as the standard unit of measure in nutrition.

12. How one can classify nutrients? Is there any food which is very essential?

Nutrients can be classified as essential or nonessential.

- Nonessential nutrients are manufactured in the body and do not need to be obtained from food. Examples include cholesterol, a fatlike substance present in all animal cells.
- Essential nutrients must be obtained from food sources, because the body either does not produce them or produces them in amounts too small to maintain growth and health. Essential nutrients include water, carbohydrates, proteins, fats, vitamins, and minerals.
- An individual needs varying amounts of each essential nutrient, depending upon such factors as gender and age. Specific health conditions, such as pregnancy, breast-feeding, illness, or drug use, make unusual demands on the body and increase its need for nutrients.

Essential Nutrients

13. The Most Important Nutrient is water: If the importance of a nutrient is judged by how long we can do without it, water ranks as the most important. A person can survive only eight to ten days without water, whereas it takes weeks or even months to die from a lack of food.

- Water circulates through our blood and lymphatic system, transporting oxygen and nutrients to cells and removing wastes through urine and sweat.
- Water also maintains the natural balance between dissolved salts and water inside and outside of cells. Our joints and soft tissues depend on the cushioning that water provides for them.
- While water has no caloric value and therefore is not an energy source, without it in our diets we could not digest or absorb the foods we eat or eliminate the body's digestive was
- The human body is 65 percent water, and it takes an average of eight to ten cups to replenish the water our bodies lose each day.

14. Sources of Water: Many foods are also a good source of water-fruits and vegetables, for instance, are 80 to 95 percent water; meats are made up of 50 percent water; and grains, such as oats and rice, can have as much as 35 percent water.

15. Some Important Points

- **Basal Metabolic Rate (B.M.R.):** It is the minimum energy requirement for maintenance of body as during rest or sleep. For a normal human adult it is 1600 kcal/day. For other animals, B.M.R. per unit weight decreases with increase in size and *vice-versa, i.e.,* a rat will have higher B.M.R. than lion.
- **Routine Metabolic Rate (R.M.R.):** It is the energy requirement of a moderately active person. R.M.R. is 2800 kcal for adult males and 2200 kcal for adult females.
- **Active Metabolic Rate (A.M.R.):** It is the energy requirement of a person doing heavy physical work. A.M.R. is 4000 - 6000 kcal depending upon the duration and intensity of work.
- **Energy Consumption/hr/kg Wt:** 1kcal during sleep, 3 kcal slow walking, 4.5 kcal during moderate cycling, 6 kcal during playing, 10 kcal during moderate running and 20.5 kcal during fast running and swimming.
- **Fuel Value:** Amount of energy liberated during complete combustion of 1 gm of a substance in bomb calorimeter. Per gram caloric value of fat is 9.3 kcal, protein 4.35 kcal and Carbohydrate 4.1 Kcal.
- **Physiological Fuel Value**: Amount of energy made available during complete oxidation of 1 gm of a substance inside the body. The value per gram is fat 9 kcal, protein 4 kcal, carbohydrate 4 kcal.
- **Balanced Diet:** It is the diet that contains all the components in optimum proportions and quantity required for maintaining the body in perfect state of health, activity and development. Various components of balanced diet are carbohydrates (60%), fats (25%), proteins (15%), vitamins (traces), minerals (traces) and roughage.

16. Carbohydrates: Carbohydrates are the human body's key source of energy, providing **4 calories of energy per gram**. When carbohydrates are broken down by the body, the sugar glucose is produced; glucose is critical to help maintain tissue protein, metabolize fat, and fuel the central nervous system.

- Glucose is absorbed into the bloodstream through the intestinal wall. Some of this glucose goes straight to work in our brain cells and red blood cells, while the rest makes its way to the liver and muscles, where it is stored as glycogen (animal starch), and to fat cells, where it is stored as fat.

Glycogen is the body's auxiliary energy source, tapped and converted back into glucose when we need more energy. Although stored fat can also serve as a backup source of energy, it is never converted into glucose. Fructose and galactose, other sugar products resulting from the breakdown of carbohydrates, go straight to the liver, where they are converted into glucose.

- Starches and sugars are the major carbohydrates. Common starch foods include whole-grain breads and cereals, pasta, corn, beans, peas, and potatoes. Naturally occurring sugars are found in fruits and many vegetables; milk products; and honey, maple sugar, and sugar cane. Foods that contain starches and naturally occurring sugars are referred to as complex carbohydrates, because their molecular complexity requires our bodies to break them down into a simpler form to obtain the much-needed fuel, glucose. Our bodies digest and absorb complex carbohydrates at a rate that helps maintain the healthful levels of glucose already in the blood.

- In contrast, simple sugars, refined from naturally occurring sugars and added to processed foods, require little digestion and are quickly absorbed by the body, triggering an unhealthy chain of events. The body's rapid absorption of simple sugars elevates the levels of glucose in the blood, which triggers the release of the hormone insulin. Insulin reins in the body's rising glucose levels, but at a price: Glucose levels may fall so low within one to two hours after eating foods high in simple sugars, such as candy, that the body responds by releasing chemicals known as anti-insulin hormones. This surge in chemicals, the aftermath of eating a candy bar, can leave a person feeling irritable and nervous.

- Many processed foods not only contain high levels of added simple sugars, they also tend to be high in fat and lacking in the vitamins and minerals found naturally in complex carbohydrates. Nutritionists often refer to such processed foods as junk foods and say that they provide only empty calories, meaning they are loaded with calories from sugars and fats but lack the essential nutrients our bodies need.

- In addition to starches and sugars, complex carbohydrates contain indigestible dietary fibers. Although such fibers provide no energy or building materials, they play a vital role in our health. Found only in plants, dietary fiber is classified as soluble or insoluble. Soluble fiber, found in such foods as oats, barley, beans, peas, apples, strawberries, and citrus fruits, mixes with food in the stomach and prevents or reduces the absorption by the small intestine of potentially dangerous substances from food. Soluble fiber also binds dietary cholesterol and carries it out of the body, thus preventing it from entering the bloodstream where it can accumulate in the inner walls of arteries and set the stage for high blood pressure, heart disease, and strokes.

- Insoluble fiber, found in vegetables, whole-grain products, and bran, provides roughage that speeds the elimination of feces, which decreases the time that the body is exposed to harmful substances, possibly reducing the risk of colon cancer. Studies of populations with fiber-rich diets, such as Africans and Asians, show that these populations have less risk of colon cancer compared to those who eat low-fiber diets, such as Americans. In the United States, colon cancer is the third most common cancer for both men and women, but experts believe that, with a proper diet, it is one of the most preventable types of cancer.

- Nutritionists caution that most Americans need to eat more complex carbohydrates. In the typical American diet, only 40 to 50 percent of total calories come from carbohydrates—a lower percentage than found in most of the world. To make matters worse, half of the carbohydrate calories consumed by the typical American come from processed foods filled with simple sugars.

- Experts recommend that these foods make up no more that 10 percent of our diet, because these foods offer no nutritional value. Foods rich in complex carbohydrates, which provide vitamins, minerals, some protein, and dietary fiber and are an abundant energy source, should make up roughly 50 percent of our daily calories.

- Dietary proteins are powerful compounds that build and repair body tissues, from hair and fingernails to muscles. In addition to maintaining the body's structure, proteins speed up chemical reactions in the body, serve as chemical messengers, fight infection, and transport oxygen from the lungs to the body's tissues. Although protein provides 4 calories of energy per gram, the body uses protein for energy only if carbohydrate and fat intake is insufficient. When tapped as an energy source, protein is diverted from the many critical functions it performs for our bodies.

17. Proteins: Proteins are made of smaller units called amino acids. Of the more than 20 amino acids our bodies require, eight (nine in some older adults and young children) cannot be made by the body in sufficient quantities to maintain health. These amino acids are considered essential and must be obtained from food. When we eat food high in proteins, the digestive tract breaks this dietary protein into amino acids. Absorbed into the bloodstream and sent to the cells that need them, amino acids then recombine into the functional proteins our bodies need.

- The amino acids which cannot be formed by humans from other amino acids are called **essential amino acids**. They are eight in number - *methionine, threonine, tryptophan, valine, leucine, isoleucine, lysine and phenylalanine.* Two amino acids are slow to be formed. They are known as semi-indispensable, *e.g., arginine, histidine*.

- Animal proteins, found in such food as eggs, milk, meat, fish, and poultry, are considered complete proteins because they contain all of the essential amino acids our bodies need.

- Plant proteins, found in vegetables, grains, and beans, lack one or more of the essential amino acids. However, plant proteins can be combined in the diet to provide all of the essential amino acids. A good example is rice lack lysine and beans lack methionine. Each of these foods lacks one or more essential amino acids, but the amino acids missing in rice are found in the beans, and vice versa. So when eaten together, these foods provide a complete source of protein.

- Experts recommend that protein intake make up only 10 percent of our daily calorie intake. Some people, especially in the United States and other developed countries, consume more protein than the body needs. Because extra amino acids cannot be stored for later use, the body destroys these amino acids and excretes their by-products.

- Alternatively, deficiencies in protein consumption, seen in the diets of people in some developing nations, may result in health problems. Marasmus and kwashiorkor, both life-threatening conditions, are the two most common forms of protein malnutrition. Some health conditions, such as illness, stress, and pregnancy and breast-feeding in women, place an enormous demand on the body as it builds tissue or fights infection, and these conditions require an increase in protein consumption. For example, a healthy woman normally needs 45 grams of protein each day.

- Experts recommend that a pregnant woman consume 55 grams of protein per day, and that a breast-feeding mother consume 65 grams to maintain health. A man of average size should eat 57 grams of protein daily. To support their rapid development, infants and young children require relatively more protein than do adults. A three-monthold infant requires about 13 grams of protein daily, and a four-year-old child requires about 22 grams. Once in adolescence, sex hormone differences cause boys to develop more muscle and bone than girls; as a result, the protein needs of adolescent boys are higher than those of girls.

18. Fats: Fats, which provide **9 calories of energy per gram**, are the most concentrated of the energy-producing nutrients, so our bodies need only very small amounts. Fats play an important role in building the membranes that surround our cells and in helping blood to clot. Once digested and absorbed, fats help the body absorb certain vitamins. Fat stored in the body cushions vital organs and protects us from extreme cold and heat. Fat consists of fatty acids attached to a substance called glycerol.

- Dietary fats are classified as saturated, monounsaturated, and polyunsaturated according to the structure of their fatty acids. Animal fats-from eggs, dairy products, and meats-are high in saturated fats and cholesterol, a chemical substance found in all animal fat.
- Vegetable fats-found, for example, in avocados, olives, some nuts, and certain vegetable oils-are rich in monounsaturated and polyunsaturated fat. As we will see, high intake of saturated fats can be unhealthy.

- To understand the problem with eating too much saturated fat, we must examine its relationship to cholesterol. High levels of cholesterol in the blood have been linked to the development of heart disease, strokes, and other health problems.

- Despite its bad reputation, our bodies need cholesterol, which is used to build cell membranes, to protect nerve fibers, and to produce vitamin D and some hormones, chemical messengers that help coordinate the body's functions. We just do not need cholesterol in our diet. The liver, and to a lesser extent the small intestine, manufacture all the cholesterol we require. When we eat cholesterol from foods that contain saturated fatty acids, we increase the level of a cholesterol-carrying substance in our blood that harms our health.

- Human diet should have more unsaturated fat because some unsaturated fatty acids cannot be synthesised from others. They are called **essential fatty acids**, *e.g., linoleic acid, linolenic acid, arachidonic acid*. Excess of lipids causes obesity, blood pressure and a number of cardiac problems. Daily requirement of adult is 50 gm.

VITAMINS

19. Both vitamins and minerals are needed by the body in very small amounts to trigger the thousands of chemical reactions necessary to maintain good health. Many of these chemical reactions are linked, with one triggering another. If there is a missing or deficient vitamin or mineral-or link anywhere in this chain, this process may break down, with potentially devastating health effects. Although similar in supporting critical functions in the human body, vitamins and minerals have key differences. Among their many functions, vitamins enhance the body's use of carbohydrates, proteins, and fats. They are critical in the formation of blood cells, hormones, nervous system chemicals known as neurotransmitters, and the genetic material deoxyribonucleic acid (DNA).

- Vitamins are classified into two groups: fat soluble and water soluble.

- Fat-soluble vitamins, which include vitamins A, D, E, and K, are usually absorbed with the help of foods that contain fat. Fat containing these vitamins is broken down by bile, a liquid released by the liver, and the body then absorbs the breakdown products and and unsaturated fats has also been associated with greater risk of developing cancers of the colon, prostate, breast, and uterus. Choosing a diet that is low in fat and cholesterol is critical to maintaining health and reducing the risk of life-threatening disease. Excess amounts of fat-soluble vitamins are stored in the body's fat, liver, and kidneys. Because these vitamins can be stored in the body, they do not need to be consumed every day to meet the body's needs.

- Water-soluble vitamins, which include vitamins C (also known as ascorbic acid), B1 (thiamine), B2 (riboflavin), B3 (niacin), B6, B12, and folic acid, cannot be stored and rapidly leave the body in urine if taken in greater quantities than the body can use. Foods that contain water-soluble vitamins need to be eaten daily to replenish the body's needs.

- In addition to the roles noted in the vitamin and mineral chart accompanying this article, vitamins A (in the form of betacarotene), C, and E function as antioxidants, which are vital in countering the potential harm of chemicals known as free radicals. If these chemicals remain unchecked they can make cells more vulnerable to cancer-causing substances. Free radicals can also transform chemicals in the body into cancer-causing

Vitamin	Function	Deficiency	Source
A (Retinol) Yellow viscous oil	Antixerophthalmic, synthesis of retinal pigments, maintenance of epithelia (prevention of keratinisation), anti-infective.	Xerophthalmia, Night blindness, keratomalacia (dermatosis, toadskin), stunted growth.	Carotene in vegetables, mango, orange and as such in milk, butter, ghee, egg yolk, liver oil.
D (Sterol-Calciferol) Ten types, common D_2 and D_3	Antirachitic, absorption and retention of Ca and P.	Rickets in children, Osteomalacia in adults.	Sunshine
E (focopherol) Colourlcss oil, α and β	Anti-sterility factor, Anti-oxidative for membrane Lipids, skin (reduces keratinisation) and hair, reduces atherosclerosis.	Erythrocyte breakdown (anaemia), muscular dysrophy (cramps), miscarriage and reduced fertility.	Green vegetables, oils, milk, cheese, butter, egg.
K (Phylloquinone) Yellow oil and fat, K_1 and K_2	Anti-haemorrhagic factor, factor for synthesis of pro-thrombin (hence essential for blood coagulation), electron carrier, liver functioning.	Prolonged bleeding (bleeding disease) or haemorrhages, liver malfunctioning.	Green vegetables, Tomato, colon bacteria.
Thiamine (B_1) vitamin G	Antineuritic, Antiberiberi, component of TPP, functioning of muscles & nerves.	Anorexia, muscular atrophy, Beri-beri in humans and polyneuritis in animals.	Yeast, whole grains, egg yolk, liver.
Riboflavin (B_2)	Maintenance of skin and oral mucosa, component of FMN and FAD, anti-migraine.	Cheilosis, glossitis, intolerance to light.	Yeast, green vegetables, milk, curd, cheese, egg.
Pantothenic Acid (B_3, alsoB_5)	Formation of CoA, required in cell respiration, functioning of skin and nerves.	Gastro-intestinal disorders, graying of hair in animals, burning feet syndrome.	Yeast, milk, wheat bran, pulses, liver. Colon bacteria.
Nicotinamide (Niacin, NicotinicAcid) N6, also B_5, B_7 Vitamin pop (pellagra preventive)	Antipellagra, component of NAD +, NADP+, energy transfer and nervous system.	Pellagra	Yeast, pulses, green vegetables, cereals, milk, egg yolk, liver.
Pyridoxine (B_6)	Component of coenzyme Pvridoxal Phosphate	Anaemia, neuritic pain, convulsions, skin lesions	Yeast, green vegetables, egg yolk, wheat germ, liver.
Biotin (B_4) Vitamin H	Prosthetic group of some enzymes in energy liberation and fat synthesis. Avidin of rag egg white prevents biotin absorption	Dermatitits, anorexia, Muscular pain, rise in blood cholesterol	Yeast, vegetables, fruits, egg yolk, colon bacteria
Folic Acid Vitamin M Pteroglutamic acid	Required in erythrocyte maturation. DNA and RNA synthesis, transfer of methyl groups.	Megaloblastic anemia, gastrointestinal disorders, glossitis	Leafy vegetables, egg yolk colon bacteria
Cobaltamine (B_{12})	Required in erythrocyte maturation. DNA and RNA synthesis	Perinicious anemia, glossitis, loss of peripheral sensation	Milk, egg, liver, fish, colon bacteria
C (Ascorbic Acid) Heat labile	Absorptoin of iron, formation of dentine, collagen and bone matrix aminio acid oxidation	Scurvy, Stiff sore joints, fragile blood vessels and bones, delayed wound healing	Citrus fruits, Guava, Tomato, Amla, Mango

Non Essential Nutrients

20. Cholestrol: Cholesterol, like fat, is a lipid-an organic compound that is not soluble in water. In order to travel through blood, cholesterol therefore must be transported through the body in special carriers, called lipoproteins. High-density lipoproteins (HDLs) remove cholesterol from the walls of arteries, return it to the liver, and help the liver excrete it as bile, a liquid acid essential to fat digestion. For this reason, HDL is called "good" cholesterol. Low-density lipoproteins (LDLs) and very-low-density lipoproteins (VLDLs) are considered "bad" cholesterol. Both LDLs and VLDLs transport cholesterol from the liver to the cells. As they work, LDLs and VLDLs leave plaque-forming cholesterol in the walls of the arteries, clogging the artery walls and setting the stage for heart disease. Almost 70 percent of the cholesterol in our bodies is carried by LDLs and VLDLs, and the remainder is transported by HDLs. For this reason, we need to consume dietary fats that increase our HDLs and decrease our LDL and VLDL levels.

21. Saturated fatty acids-found in foods ranging from beef to ice cream, to mozzarella cheese to doughnuts-should make up no more than 10 percent of a person's total calorie intake each day. Saturated fats are considered harmful to the heart and blood vessels because they are thought to increase the level of LDLs and VLDLs and decrease the levels of HDLs. Monounsaturated fats-found in olive, canola, and peanut oils appear to have the best effect on blood cholesterol, decreasing the level of LDLs and VLDLs and increasing the level of HDLs. Polyunsaturated fats-found in margarine and sunflower, soybean, corn, and safflower oils-are considered more healthful than saturated fats. However, if consumed in excess (more than 10 percent of daily calories), they can decrease the blood levels of HDLs.

NUTRITIONAL DISORDERS:

21. Marasmus: It is a disorder caused by deficiency of food or protein energy malnutrition (PEM). Marasmus occurs in infants (1-2 years) when they do not get sufficient mother's milk and weaning diet. Such infants are often provided with much diluted cow or buffalo milk. Marasmus babies have shriveled appearance with thin face, sunken eyes, wrinkled and dry skin, protruding ribs, reduced digestion and frequent diarrhea, low weight, little growth and poor mental abilities.

22. Kwashiorkor. It is a protein energy malnutrition that occurs in children of 1-4 years due to constant reduced protein in diet. This can be due to (i) early stoppage of breast milk (ii) late supply of supplementary food (iii) over diluted milk (iv*)* maintaining children on cooking water of cereals. A kwashiorkor child has match stick legs, protruded belly, blotchy skin, anaemia, oedema of certain parts and repeated diarrhoea.

23. Obesity: Obesity is a luxury nutritional disorder caused by greater intake of food than the requirement of body. It is quite common in persons having higher intake of sweets, carbohydrate rich food, fried articles, fat rich food and absence of roughage in food. Obese persons are overweight with excessive accumulation of fat. Movements are reduced. There is a risk of diabetes, osteoarthritis, gall bladder stones, high blood pressure and cardiac problems. The reasons for overeating are psychogenic, physiological, genetic and developmental.

24. Hyper-cholesterolemia: It is characterized by high blood cholesterol due to higher intake of cholesterol rich eggs, red meat, butter and ghee or reduced biotin intake. Cholesterol gets deposited on the walls of blood vessels making their lumen narrow and walls stiffer. This increases blood pressure or hypertension. It leads to other cardio-vascular disorders.

25. Rickets: The disorder is due to deficient ossification of bone ends leading to their bending and swelling of joints (knee, wrist, elbow). It occurs in children of 6 months to 2 years when their diet is deficient in vitamin D, calcium or phosphorus.

26. Osteomalacia: The disorder occurs in adults when their food is deficient in vitamin D, calcium or phosphorus. Bones of vertebral column become weak. Pelvic and other bones bend and become soft.

27. Xerophthalmia: It is a vitamin A deficiency disorder which is caused by stoppage of lacrimal activity. Cornea and eye lids become dry and ulcerated. It leads to swelling of dry eye lids and opacity of cornea. Xerophthalmia, if untreated, leads to blindness.

28. Night Blindness (Nyctalopia): Dietary deficiency of vitamin A results in deficiency of rhodopsin or visual purple. The latter impairs the ability to see in dim light or night.

29. Beri-Beri: Beriberi is a debility produced by deficiency of thiamine as during excessive intake of polished rice. It results in loss of appetite (anorexia), weakening of muscles, fatigue, poor digestion, palpitation, slow reflexes, body rigidity and neuritis. There is inflammation of peripheral nerves, weakening of muscles, paralysis and progressive oedema. Depending upon the system affected, beri-beri is of five types - dry, wet, mixed, cardiac and cerebral.

30. Scurvy (Sailor's Disease): Vitamin C (ascorbic acid) deficiency leads to scurvy characterized by bleeding gums, falling of teeth, fragile blood vessels, fragile bones and slow healing of injuries.

31. Anaemia. It is disorder in which haemoglobin content of the blood is low due to either few red blood corpuscles or their haemoglobin content. It is of three types

- **Microcytic** (most common) due to iron deficiency leading to fewer and smaller erythrocytes with reduced haemoglobin.
- **Megablastic** due to fewer abnormal' red blood cells caused by deficiency of folic acid or B_{12}.
- **Pernicious** due to production of haemoglobin free immature RBCs caused by B_{12} deficiency as a result of deficient intrinsic factor. Sickle cell anemia and thalassemia are genetically controlled. Anemia also occurs due to excessive blood loss. The same is called post-haemorrhagic anemia.

32. MINERALS:

- **Sodium.** It is main cation of extracellular fluid, component of bile salts which is involved in osmotic balance, acid-basic balance, absorption of glucose, electro-chemical impulse conduction in nerves and muscles. Deficiency produces cramps.
- **Chlorine:** Main anion of extracellular fluid, involved in HCl synthesis and acid-basic balance.
- **Potassium:** The cation is present in intracellular as well as extracellular fluid. It takes part in muscle and nerve activity, glycogen and protein synthesis.
- **Magnesium:** It is an enzyme activator, component of bones and teeth besides being required for muscle relaxation. Deficiency produces convulsions.
- **Sulphur:** It is a constituent of many proteins, enzymes and coenzymes.
- **Cobalt:** A component of B_{12}. Deficiency causes

pernicious anaemia.

- **Fluorine:** Maintains enamel and checks dental decay. In excess harmful to teeth and bones.
- **Calcium:** It is the major component of bones and teeth. Calcium is also required for blood clotting, muscle contraction, A TP-ase, nerve impulse transmission, heart functioning, etc. Vitamin D is essential for calcium absorption. Deficiency produces rickets and muscle spasms.
- **Phosphorus:** Along with calcium, it occurs in bones and teeth. Phosphorus is a component of nucleic acids, phospholipids, A TP and some coenzymes. Phosphorylation activates sugars and fatty acids. Phosphate is *pH* buffer. Deficiency reduces growth, metabolism and causes rickets in children (along with calcium and vitamin D).
- **Iron:** It is re-used in the body. Iron is a component of haemoglobin, myoglobin, cytochromes and ferredoxin. Deficiency leads to anaemia.
- **Iodine:** It is essential for production of hormone thyroxine of thyroid glands. Deficiency causes goitre.
- **Selenium:** Essential for protecting sperms and hence maintenance of male fertility.

C. Stablizing Interactions

Covalent bonds are very stable because the energies required to break them are much greater than the thermal energy available at room temperature (25ºC) or body temperature (37ºC). For example, the thermal energy at 25 ºC is approximately 0.6 kilocalorie per mole (kcal/mol), whereas the energy required to break the carbon-carbon single bond (C-C) in ethane is about 140 times larger. Consequently at room temperature (25 ºC), fewer than 1 in 10^{12} ethane molecules is broken into a pair of $\cdot CH_3$ radicals, each containing an unpaired, nonbonding electron.

Covalent single bonds in biological molecules have energies similar to that of the C-C bond in ethane. Because more electrons are shared between atoms in double bonds, they require more energy to break than single bonds. For instance, it takes 84 kcal/mol to break a single C-C bond, but 170 kcal/mol to break a C=O double bond. The most common double bonds in biological molecules are C=O, C=N, C=C, and P=O.

The energy required to break noncovalent interactions is only 1–5 kcal/mol, much less than the bond energies of covalent bonds. Indeed, noncovalent interactions are weak enough that they are constantly being formed and broken at room temperature. Although these interactions are weak and have a transient existence at physiological temperatures (25–37 ºC), multiple noncovalent interactions can act together to produce highly stable and specific associations between different parts of a large molecule or between different macromolecules. We first review the four main types of noncovalent interactions and then consider their role in the binding of biomolecules to one another and to other molecules.

described it. These nonspecific interactions result from the momentary random fluctuations in the distribution of the electrons of any atom, which give rise to a transient unequal distribution of electrons, that is, a transient electric dipole. If two noncovalently bonded atoms are close enough together, the transient dipole in one atom will perturb the electron cloud of the other. This perturbation generates a transient dipole in the second atom, and the two dipoles will attract each other weakly. Similarly, a polar covalent bond in one molecule will attract an oppositely oriented dipole in another.

Van der Waals interactions, involving either transient induced or permanent electric dipoles, occur in all types of molecules, both polar and nonpolar. In particular, van der Waals interactions are responsible for the cohesion between molecules of nonpolar liquids and solids, such as heptane, $CH_3—(CH_2)_5—CH_3$, that cannot form hydrogen bonds or ionic interactions with other molecules. When these stronger interactions are present, they override most of the influence of van der Waals interactions. Heptane, however, would be a gas if van der Waals interactions could not form.

The strength of van der Waals interactions decreases rapidly with increasing distance; thus these noncovalent bonds can form only when atoms are quite close to one another. However, if atoms get too close together, they become repelled by the negative charges in their outer electron shells. When the van der Waals attraction between two atoms exactly balances the repulsion between their two electron clouds, the atoms are said to be in van der Waals contact.

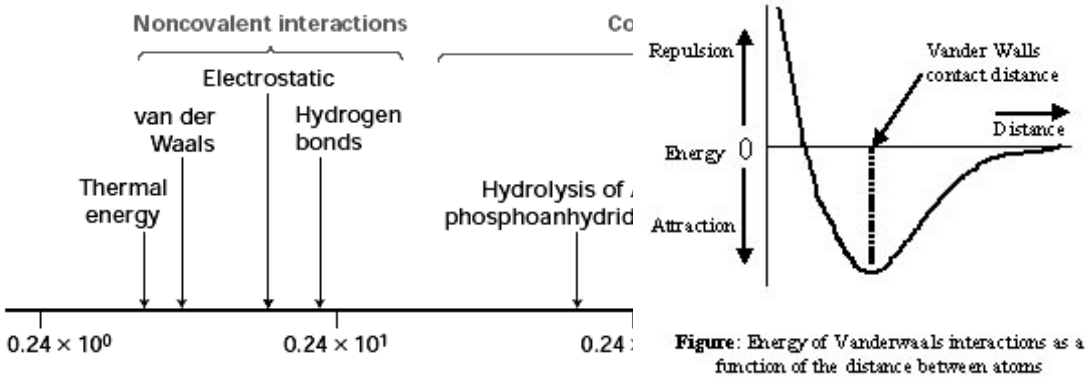


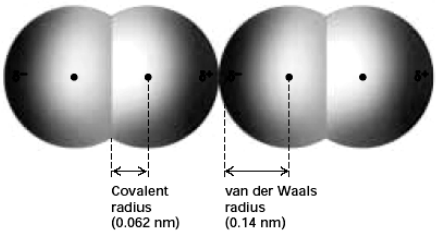
Figure: Energy of Vanderwaals interactions as a function of the distance between atoms

C1. Van Der Waal's Interactions:

When any two atoms approach each other closely, they create a weak, nonspecific attractive force that produces a van der Waals interaction, named for Dutch physicist Johannes Diderik van der Waals (1837 –1923), who first

Each type of atom has a van der Waals radius at which it is in van der Waals contact with other atoms. The van der Waals radius of an H atom is 0.1 nm, and the radii of O, N, C, and S atoms are between 0.14 and 0.18 nm. Two covalently bonded atoms are closer together than two atoms that are merely in van der Waals contact. For a van der Waals

interaction, the internuclear distance is approximately the sum of the corresponding radii for the two participating atoms. Thus the distance between a C atom and an H atom in van der Waals contact is 0.27 nm, and between two C atoms is 0.34 nm. In general, the van der Waals radius of an atom is about twice as long as its covalent radius. For example, a C—H covalent bond is about 0.107 nm long and a C—C covalent bond is about 0.154 nm long.



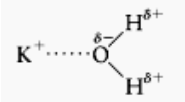
The energy of the van der Waals interaction is about 1 kcal/mol, only slightly higher than the average thermal energy of molecules at 25°C. Van der Waals interactions, as well as other noncovalent bonds, mediate the binding of many enzymes with their specific substrates (the substances on which an enzyme acts) and of each type of antibody with its specific antigen.

C2. Electrostatic (Ionic) Interactions:

In some compounds, the bonded atoms are so different in electronegativity that the bonding electrons are never shared: these electrons are always found around the more electronegative atom. In sodium chloride (NaCl), for example, the bonding electron contributed by the sodium atom is completely transferred to the chlorine atom. Even in solid crystals of NaCl, the sodium and chlorine atoms are ionized, so it is more accurate to write the formula for the compound as Na^+Cl^-.

Because the electrons are not shared, the bonds in such compounds cannot be considered covalent. They are, rather, ionic bonds (or interactions) that result from the attraction of a positively charged ion —a cation —for a negatively charged ion —an anion. Unlike covalent or hydrogen bonds, ionic bonds do not have fixed or specific geometric orientations because the electrostatic field around an ion —its attraction for an opposite charge —is uniform in all directions. However, crystals of salts such as Na^+Cl^- do have very regular structures because that is the energetically most favorable way of packing together positive and negative ions. The force that stabilizes ionic crystals is called the lattice energy.

In aqueous solutions, simple ions of biological significance, such as Na^+, K^+, Ca^{2+}, Mg^{2+}, and Cl^-, do not exist as free, isolated entities. Instead, each is surrounded by a stable, tightly held shell of water molecules. An ionic interaction occurs between the ion and the oppositely charged end of the water dipole, as shown below for the K+ ion:

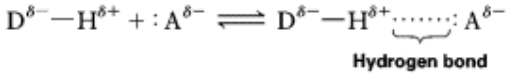


Ions play an important biological role when they pass through narrow, protein-lined pores, or channels, in membranes. For example, ionic movements through membranes are essential for the conduction of nerve impulses and for the stimulation of muscle contraction. The ions must lose their shell of water molecules in order to pass through ion channel proteins; channel proteins can then selectively admit only Na^+, or K^+, or Ca^{2+} ions, a selectivity essential for nerve function.

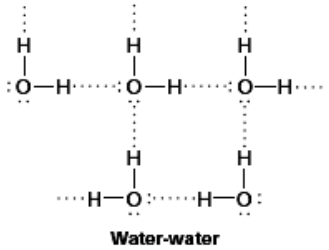
Most ionic compounds are quite soluble in water because a large amount of energy is released when ions tightly bind water molecules. This is known as the energy of hydration. Oppositely charged ions are shielded from one another by the water and tend not to recombine. Salts like Na^+Cl^- dissolve in water because the energy of hydration is greater than the lattice energy that stabilizes the crystal structure. In contrast, certain salts, such as $Ca_3(PO_4)_2$, are virtually insoluble in water; the large charges on the Ca^{2+} and PO_4^{3-} ions generate a formidable lattice energy that is greater than the energy of hydration.

C3. The Hydrogen Bond:

Normally, a hydrogen atom forms a covalent bond with only one other atom. However, a hydrogen atom covalently bonded to a donor atom, D, may form an additional weak association, the hydrogen bond, with an acceptor atom, A:



In order for a hydrogen bond to form, the donor atom must be electronegative, so that the covalent D—H bond is polar. The acceptor atom also must be electronegative, and its outer shell must have at least one nonbonding pair of electrons that attracts the d+ charge of the hydrogen atom. In biological systems, both donors and acceptors are usually nitrogen or oxygen atoms, especially those atoms in amino (—NH_2) and hydroxyl (—OH) groups. Because all covalent N—H and O—H bonds are polar, their H atoms can participate in hydrogen bonds. By contrast, C—H bonds are nonpolar, so these H atoms are almost never involved in a hydrogen bond.



Water molecules provide a classic example of hydrogen bonding. The hydrogen atom in one water molecule is attracted to a pair of electrons in the outer shell of an oxygen atom in an adjacent molecule. In general, molecules with polar bonds that easily form hydrogen bonds with water can dissolve in water and are said to be hydrophilic (Greek, "water-loving"). Besides the hydroxyl and amino groups, peptide and ester bonds are important chemical groups that interact well with water:

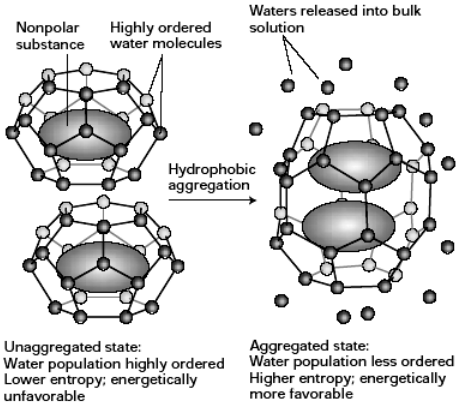
Most hydrogen bonds are 0.26 –0.31 nm long, about twice the length of covalent bonds between the same atoms. In particular, the distance between the nuclei of the hydrogen and oxygen atoms of adjacent hydrogen-bonded molecules in water is approximately 0.27 nm, about twice the length of the covalent O—H bonds in water. The hydrogen atom is closer to the donor atom, D, to which it remains covalently bonded, than it is to the acceptor. The length of the covalent D—H bond is a bit longer than it would be if there were no hydrogen bond, because the acceptor "pulls" the hydrogen away from the donor. The strength of a hydrogen bond in water (≈5 kcal/mol) is much weaker than a covalent O—H bond (≈110 kcal/mol).

Hydrogen Bonds is a major Stabilizing Force in Macromolecules

An important feature of all hydrogen bonds is directionality. In the strongest hydrogen bonds, the donor atom, the hydrogen atom, and the acceptor atom all lie in a straight line. Nonlinear hydrogen bonds are weaker than linear ones. The strengths of the hydrogen bonds in proteins and nucleic acids are only 1 to 2 kcal/mol, considerably weaker than the hydrogen bonds between water molecules. Initially, both the —OH and —NH2 groups in the protein are hydrogen-bonded to water, and the formation of a hydrogen bond between these groups involves disruption of their hydrogen bonds with water. Thus the net change in energy in forming this —OH···N hydrogen bond will be less than the 5 kcal/mol characteristic of hydrogen bonds between water molecules.

C4. Hydrophobic Interactions:

Nonpolar molecules do not contain ions, possess a dipole moment, or become hydrated. Because such molecules are insoluble or almost insoluble in water, they are said to be hydrophobic (Greek, "water-fearing"). The covalent bonds between two carbon atoms and between carbon and hydrogen atoms are the most common nonpolar bonds in biological systems. Hydrocarbons —molecules made up only of carbon and hydrogen —are virtually insoluble in water. A large triacylglycerol (or triglyceride) such as tristearin, a component of animal fat, is also insoluble in water, even though its six oxygen atoms participate in some slightly polar bonds with adjacent carbon atoms. When shaken in water, tristearin forms a separate phase similar to the separation of oil from the water-based vinegar in an oil-and-vinegar salad dressing.



The force that causes hydrophobic molecules or nonpolar portions of molecules to aggregate together rather than to dissolve in water is called the hydrophobic bond. This is not a separate bonding force; rather, it is the result of the energy required to insert a nonpolar molecule into water. A nonpolar molecule cannot form hydrogen bonds with water molecules, so it distorts the usual water structure, forcing the water into a rigid cage of hydrogen-bonded molecules around it. Water molecules are normally in constant motion, and the formation of such cages restricts the motion of a number of water molecules; the effect is to increase the structural organization of water. This situation is energetically unfavorable because it decreases the randomness (entropy) of the population of water molecules. The opposition of water molecules to having their motion restricted by forming cages around hydrophobic molecules or portions thereof is the major reason molecules such as tristearin and heptane are essentially insoluble in water and interact mainly with other hydrophobic molecules. Nonpolar molecules can also bond together, albeit weakly, through van der Waals interactions. The net result of the hydrophobic and van der Waals interactions is a very powerful tendency for hydrophobic molecules to interact with one another, and not with water.

The order of hydrophobic character for various amino acids is as follows

Hydrophobic:
Glycine<Tyorine<Alanine<Histidine<Methione<Valine<leucine<Isolecine<Phenyl alanine<Tryptophan

Small hydrocarbons like butane (CH_3—CH_2—CH_2—CH_3) are somewhat soluble in water, because they can dissolve without disrupting the water lattice appreciably. However, 1-butanol (CH_3—CH_2—CH_2—CH_2OH) mixes completely with water in all proportions. The replacement of just one hydrogen atom with the polar —OH group allows the molecule to form hydrogen bonds with water and greatly increases its solubility. Simply put, like dissolves like. Polar molecules dissolve in polar solvents such as water, while nonpolar molecules dissolve in nonpolar solvents such as hexane

Molecular Complementarity Permits Tight, Highly Specific Binding of Biomolecules

Both inside and outside of cells, ions and molecules are constantly bumping into one another. The greater the number of copies of any two types of molecules per unit volume (i.e., the higher their concentration), the more likely they are to encounter one another. When two molecules encounter each other, they most likely will simply bounce apart because the noncovalent interactions that would bind them together are weak and have a transient existence at physiological temperatures. However, molecules that exhibit molecular complementarity, a lock-and-key kind of fit between their shapes, charges, or other physical properties, can form multiple noncovalent interactions at close range. When two such structurally complementary molecules bump into each other, they can bind (stick) together.

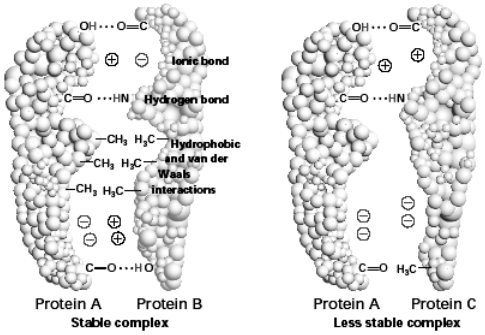


Figure illustrates how multiple, different weak bonds can bind two proteins together. Almost any other arrangement of the same groups on the two surfaces would not allow the molecules to bind so tightly. Such multiple, specific interactions between complementary regions within a protein molecule allow it to fold into a unique three-dimensional shape and hold the two chains of DNA together in a double helix. Similar interactions underlie the association of groups of more than two molecules into multimolecular complexes, leading to formation of muscle fibers, to the gluelike associations between cells in solid tissues, and to numerous other cellular structures.

Depending on the number and strength of the noncovalent interactions between the two molecules and on their environment, their binding may be tight (strong) or loose (weak) and, as a consequence, either long-lasting or transient. The higher the affinity of two molecules for each other, the better the molecular “fit” between them, the more

noncovalent interactions can form, and the tighter they can bind together. An important quantitative measure of affinity is the binding dissociation constant Kd.

Nearly all the chemical reactions that occur in cells also depend on the binding properties of enzymes. These proteins not only speed up reactions but also do so with a high degree of specificity, a reflection of their ability to bind tightly to only one or a few related molecules. Indeed, the binding specificity of large biological molecules, particularly proteins and nucleic acids, is one of the distinctive features that distinguish biochemistry from typical solution chemistry. Clearly, molecular complementarity and noncovalent interactions underlie the structures of biomolecules and many processes critical to life.

D. Principles of biophysical chemistry

D1. Water, pH and Buffers

Water is essential to biochemistry because:
(a) Biological macromolecules assume specific shapes in response to the chemical and physical properties of water.
(b) Biological molecules undergo chemical reactions in an aqueous environment.
(c) Water is a key reactant in many reactions, usually in the form of H^+ and OH^- ions.
(d) Water is oxidized in photosynthesis to produce molecular oxygen, O_2, as part of the process that converts the sun's energy into usable chemical form. Expenditure of that energy under aerobic conditions leads to the reduction of O_2 back to water.

Physical Properties of Water

1. The structure of water closely approximates a tetrahedron with its two hydrogen atoms and the two lone pairs of electrons of its oxygen atom "occupying" the vertices of the tetrahedron.

2. The high electronegativity of oxygen relative to hydrogen results in the establishment of a permanent dipole in water molecules.

3. The polar nature of water results in negative portions of the molecule being attracted to the positive portions of neighboring water molecules by a largely electrostatic interaction known as the hydrogen bond.

4. Hydrogen bonds are represented as D—H···A, where D—H is a weakly acidic compound so that the hydrogen atom (H) has a partial positive charge, and A is a weakly basic group that bears lone pairs of electrons. A is often an oxygen atom or a nitrogen atom (occasionally sulfur).

5. Water is strongly hydrogen bonded, with each water molecule participating in four hydrogen bonds with its neighbors; two in which it donates and two in which it accepts. Hydrogen bonds commonly form between water molecules and the polar functional groups of biomolecules, or between the polar functional groups themselves.

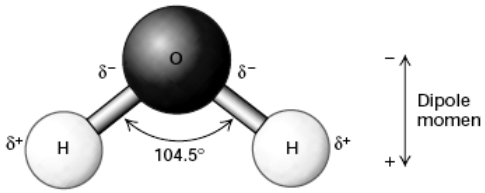
6. The strongly hydrogen bonded character of water is responsible for many of its characteristic properties, most notably:
(a) A high heat of fusion, which allows water to act as a heat sink, such that greater heat loss is necessary for the freezing of water compared to other substances of similar molecular mass.
(b) A high heat of vaporization, such that relatively more heat must be input to vaporize water compared to other substances of similar molecular mass.
(c) An ability to dissolve most polar compounds.
(d) An open structure makes ice less dense than liquid water, thereby making ice float, insulating the water beneath it, and inhibiting total freezing of large bodies of water.

7. A variety of weak electrostatic interactions are critical to the structure and reactivity of biological molecules. These interactions include, in order of increasing strength, London dispersion forces, dipole–dipole interactions, hydrogen bonds, and ionic interactions.

8. Water is an excellent solvent of polar and ionic substances due to its property of surrounding polar molecules and ions with oriented shells of water, thereby attenuating the electrostatic forces between these molecules and ions.

9. The tendency of water to minimize its contact with nonpolar (hydrophobic) molecules is called the hydrophobic effect. This effect is largely driven by the increase in entropy caused by the necessity for water to order itself around nonpolar molecules. This causes the nonpolar molecules to aggregate, thereby reducing the surface area that water must order itself about. Consequently, nonpolar substances are poorly soluble in aqueous solution.

10. Many biological molecules have both polar (or charged) and nonpolar functional groups and are therefore simultaneously hydrophilic and hydrophobic. These molecules are said to be amphiphilic or amphipathic.



11. Osmosis is the movement of solvent across a semipermeable membrane from a region of lower concentration of solute to a region of higher concentration of solute. Osmotic pressure of a solution is the pressure that must be applied to the solution to prevent an inflow of solvent. Hence, an increase in solute concentration results in an increase in osmotic pressure.

12. Diffusion is the random movement of molecules in solution (or in the gas phase). It is responsible for the movement of solutes from a region of high concentration to a region of low concentration.

Chemical Properties of Water

1. Water is a neutral, polar molecule that has a slight tendency to ionize into H+ and OH–. However, the proton is never free and binds to a water molecule to form H3O+ (hydronium ion).

2. The ionization of water is described as an equilibrium between the unionized water (reactant) and its ionized species (products)

$$H_2O \rightleftharpoons H^+ + OH^-$$

In which

$$K = \frac{[H^+][OH^-]}{[H_2O]}$$

Since in dilute aqueous solution, the concentration of water is essentially constant (55.5 M), the concentration of H2O is incorporated into the value of K, which is referred to as Kw, the ionization constant of water.

$$K_w = [H^+][OH^-]$$

3. The values of both H^+ and K are inconveniently small; hence, their values are more conveniently expressed as negative logarithms, so that

pH= -log[H^+]
pK= -log *K*

4. According to the Brønsted–Lowry definition, an acid is a substance that can donate a proton, and a base is a substance that can accept a proton. The strength of a weak acid is proportional to its dissociation constant, which is expressed as

$$K = \frac{[H^+][A^-]}{[HA]}$$

5. The pH of a solution of a weak acid is determined by the relative concentrations of the acid and its conjugate base. The equilibrium expression for the dissociation of a weak acid can be rearranged to

$$pH = pK + \log \frac{[A^-]}{[HA]}$$

This relationship is known as the Henderson–Hasselbalch equation. When the concentration of a weak acid is equal to the concentration of its conjugate base, pH = pK. Hence, the stronger the acid, the lower its pK.

6. Solutions of a weak acid at pH's near its pK resist large changes in pH as OH– or H+ is added. Added protons react with the weak acid's conjugate base to reform the weak acid; whereas added OH– combines with the acid to form its conjugate base and water. A solution of a weak acid and its conjugate base (in the form of a salt) is referred to as a buffer. Buffers are effective within 1 pH unit of the pK of the component acid.

7. The properties of most proteins, enzymes for example, are sensitive to pH. **As the pH drops**, H^+ bind to the carboxyl groups (COO^-) of aspartic acid (Asp) and glutamic acid (Glu), neutralizing their negative charge, and H^+ bind to the unoccupied pair of electrons on the N atom of the amino (NH_2) groups of lysine (Lys) and arginine (Arg) giving them a positive charge.

The result: Not only does the net charge on the molecule change (it becomes more positive) but many of the opportunities that its R groups have for ionic (electrostatic) interactions with other molecules and ions are altered.

8. As the pH rises, H^+ are removed from the COOH groups of Asp and Glu, giving them a negative charge (COO^-) , and H^+ are removed from the NH_3^+ groups of Lys and Arg removing their positive charge.

The result: Again the net charge on the molecule changes (it becomes more negative) and, again, many of the opportunities its R groups have for electrostatic interactions with other molecules or ions are altered.

9. The pH of the **cytosol** within a human cell is about 7.4. BUT, this value masks the pH differences that are found in various compartments within the cell. For example, The interior of **lysosomes** is much more acidic (as low as pH 4-5) than the cytosol, and the enzymes within work best at these low pH values.

10. The pH differential created within **chloroplasts** by the energy of the sun is harnessed to synthesize ATP which, in turn, powers the synthesis of food. The pH differential created within **mitochondria** during the respiration of food is harnessed to the synthesis of ATP which, in turn, powers most of the energy-consuming activities of the cell such as locomotion and biosynthesis of cell components.

11. A ten-fold change in [H^+] alters the pH by 1.0 unit.

12. Buffers: A buffer is a mixture of a weak acid and the salt of that acid. The practical buffer zone is pKa ± 1 pH. The Henderson-Hassalbalch equation is used for calculations involving buffers.

$$pH = pKa + \log \frac{[A^-]}{[HA]}$$

pH of a buffer is theoretically independent of dilution (depends only on the ration of A^-/HA and pKa).

13. Solving pH problems

- **Strong acids** (completely dissociated)
  [H+] = normality of acid pH = -log[H^+]
- **Strong bases** (completely dissociated)
  [OH-] = normality of base pOH = -log[OH^-]
  pH + pOH = 14
- **Weak acids**: two kinds of problems

a. % ionization x molarity = [H^+]

b. $Ka = \frac{[H^+][A^-]}{[HA]} = \frac{x^2}{[HA] - x} \sim \frac{x^2}{[HA]}$

$x = [H^+] = \{Ka \times [HA]\}$

D2. Kinetics, Dissociation and Association constants

1. The branch of physical chemistry in which we study rate of reaction and mechanism of the reaction is known as *chemical kinetics*

2. **Rate of reaction**: The rate of the reaction may be defined as, "the rate of change of the concentration of the reactants with time. In other words, it is measured by the amount of substance in unit volume changed in unit time and amount is measured in mole. Thus,

Rate of reaction = $\frac{\text{Amount transformed}}{\text{Time taken in transformation}}$

= dx /dt

Where dx is the amount of substance changed and dt is a very small interval of time. Negative sign indicates that velocity of the reaction decreases with time. It is important to note that the concept of mechanical velocity or speed cannot be used in measuring the rate of reaction. Because rate of reaction depends upon the molar concentrations of the reactants and it decreases with time, therefore rate of reaction varies with time.

3. Factors influencing rate of reaction

i. **Effect of concentration:** The rate of a reaction decreases with decrease of concentration.

ii. **Effect of temperature**: It is observed that rate of a reaction increase with rise of temperature. In general, rate of a reaction becomes double on rise of 10^0C (usually 25-35°C).

iii. **Effect of the nature of reactant:** It is observed that the reactions in which simple ions are involved; take place faster than the reactions in which ions of considerable bonds are involved. For example, oxidation of Fe^{2+} ion by permanganate ion (MnO_4^- ion) in acidic medium takes place faster than the oxidation of $C_2O_4^{2-}$ ion under same conditions.

iv. **Effect of Catalyst**: In general catalyst increases the rate of reaction.

v. **Effect of Surface area of reactants**: It is observed that the smaller particles react more rapidly than the larger particles of the same mass. For example, coal dust bums rapidly than that of large lumps of coal.

vi. **Effect of radiation:** Rate of certain reactions increases by the absorption of photons of certain radiations. Such reactions are known as photochemical reactions.

4. **Rate Constant:** According to collision theory, the rate of the reaction is proportional to the number of molecular collisions taking place per second. Thus, on increasing the concentration of the reactant number of collisions increases, hence rate of the reaction increases.Thus, for a general reaction

$$A \rightarrow \text{Products}$$
$$r = dx/dt = dCA/dt = kC_A$$

Where r = rate of reaction, C_A = Concentration of the reactant A and k = rate constant.

If $C_A = 1$, then r = k.

We can say that at a given temperature, rate is equal to the rate constant of reaction when concentration of the reactant is unity. Thus rate constant is known as *specific reaction rate.*

In the case of two reactants, the reaction may be written as

$$A + B \rightarrow \text{Products}$$
$$r = dx/dt = kC_AC_B$$

where all the terms have usual meaning. If $C_A = C_B = 1$, then r = k. Thus rate constant is equal to rate of the reaction when concentration of each of the reactants is unity.

5. Order of a reaction: Chemical reactions may be classified on the basis of the number of molecules that must ultimately react to form the reaction products. Thus we have monomolecular, bimolecular; and tri-molecular reactions, in which one, two or three molecules, respectively undergo reaction.

Chemical reactions are classified on a kinetic basis by reaction order such as zero-order, first- order, second order, and third order reactions, depending on how the reaction rate is influenced by the concentration of the reactants under a given set of conditions.

6. Zero Order Reactions: If the rate of reaction is independent of the concentration of the reactant then the reaction is called zero *order reaction.* For example

$$A \rightarrow \text{Product}$$
$$r = -dC^0_A/dt = k$$

If a = initial concentration, x = amount decomposed in time't' then

$$K\frac{a-(a-x)}{t} = \frac{x}{t}$$

7. First Order Reactions: If the rate of a reaction is determined by the variation of concentration term of only one reactant then the reaction is called first order reaction. For example,

$$A \rightarrow \text{Product}$$
$$r = dCA/dt = kC_A$$

$$-dC_A/C_A = K.dt$$

The integrated form of the first-order rate equation is

$$k = \frac{2.303}{t}\log\frac{C_0}{C}$$

Where Co and Care concentration at initial time (i.e., t = 0) and at any subsequent time respectively. If the initial concentration of the reactant is '**a**' mole L^{-1} and after time **t**, 'x' mole L^{-1} have reacted to give the product, then the concentration of the reactant left behind at the time **t** will be (a-x) moles L^{-1}. Then the first order rate equation may be given as

$$k = \frac{2.303}{t}\log\frac{a}{a-x}$$

In first order reaction, the half time t ½ of the reaction is given by:

$$t_{½} = 0.693/k$$

Here, the half-time is independent of the initial concentration of the substrate. t ½ is defined as the time during which half the concentration of a given reactants react. First order rate constants have the dimensions of reciprocal time, usually sec^{-1}

8. Second Order Reactions: A reaction is said to be of the second order if its rate is determined by the variation of two concentration terms or rate of the reaction is proportional to the second power of the concentration of a single substance.

In general, the second order reactions may be of two types.

a. When the concentration of the reactants is same i.e.

$$A + A \rightarrow \text{Products}$$

Let 'a' be the initial concentration of each of the reactants to start the reaction and x be the amount of each reactant decomposed after time, t.

2A	→	Products
a		0 (initial concentration)
(a-x)		x (concentration after time t).

The rate of the reaction may be represented by

$$\frac{dx}{(a-x)^2} = K.dt$$

On integration within the limits 0 to x at time 0 to t respectively, we get

$$K = \frac{1}{t}X\frac{x}{a(a-x)}$$

b. When the concentration of the reactants is different, i.e.

A + B	→	Product
a b		0 (initial concentration)
(a-x) (b-x)		x (concentration after time t)

The rate of reaction will be expressed as

$$\frac{dx}{dt} = k(a-x)(b-x)$$

$$\frac{dx}{(a-x)(b-x)} = K.dt$$

On intergration within the limits 0 to x at time 0 to t respectively,

$$k = \frac{2.303}{t(a-b)} \log 10 \frac{(a-x)b}{(b-x)a}$$

The rate constants of second order reactions have the dimensions 1/concentration x time or mol^{-1} sec^{-1}.

9. Third Order reactions: A reaction is said to be of third order if the rate is determined by the variation of three concentration terms. In other words, the minimum number of molecules necessary for the reaction to take place is three. There may be three different cases in third order reaction.

a. All the three species have equal concentrations

A+A+A → P

Consider a general reaction involving three moles of A having the concentration of each as 'a' moles per litre. Let the concentration at time t becomes (a-x) then the rate of reaction is given by

$$\frac{dx}{dt} = k(a-x)^3$$

The integrated form of the equation is

$$K = \frac{1}{t} \frac{2a-x}{2a^2(a-x)}$$

b. When the concentrations of two reactants are unequal and are different

2A + B → Products

The rate of reaction is given by

$$\frac{dx}{dt} = k(a-2x)^2(b-x)$$

In order to know the values of k, integrating above equation, we get

$$K = \frac{1}{t} \frac{1}{(a-2b)^2} \left[\frac{2x(2b-a)}{a(a-x)} + \log e \frac{\left(b(a-2x)\right)}{a(b-x)} \right]$$

The rate constants of third-order reactions have the, dimensions $mol^{-2}L^2sec^{-1}$

D3. Thermodynamics

1. Matter is "**anything that has mass and takes up space**". Energy is the capacity to do work - that is, to move matter against opposing forces, such -as gravity and friction.

2. Anything that moves possesses a form of energy called kinetic energy. Any moving object, including an atom or molecule, has **kinetic energy**- the energy of motion. **Heat**, or **thermal energy**, is kinetic energy that results from the random movement of molecules. **Light** also represents energy, which can be harnessed to perform work, such as powering photosynthesis in green plants.

3. A resting object, not presently at work, may also possess energy, which is the capacity to do work. Stored energy, or **potential energy**, is energy that matter possesses because of its location or arrangement. Water behind a dam, for instance, stores energy because of its altitude. Chemical energy, a form of potential energy especially important to biologists, is stored in molecules because of the arrangement of the atoms that are bonded together.

4. Energy Transformations: Energy can be converted from one form to another. Chemical energy can be tapped when chemical reactions rearrange the atoms of molecules in such a way that potential energy stored in the molecules is converted to kinetic energy. The chemical energy stored in the fuel molecules had itself been converted from light energy by plants during photosynthesis.

5. Free energy is the portion of a system's energy that can perform work when temperature is uniform throughout the system, as in a living cell. A system's quantity of free energy is symbolized by the letter "**G**". There are two components to "**G**": the system's total energy (symbolized by **H**) and its entropy (symbolized by **S**). Free energy is related to these factors in the following way:

G=H-TS

T stands for absolute temperature (^oC + 273). Notice that temperature amplifies the entropy term of the equation.

6. In any spontaneous process, the free energy of a system decreases. The change in free energy as a system goes from a starting state to a different state is represented by ΔG:

$\Delta G = G_{\text{final state}} - G_{\text{Starting state}}$

or put another way:

ΔG=ΔH -TΔS

7. The study of the energy transformations that occur in a collection of matter is called **thermodynamics**. Scientists use the word system to denote the collection of matter under study and refer to the rest of the universe - everything outside the system - as the surroundings.

8. **A closed system**, such as that approximated by liquid in a thermos bottle, is isolated from its surroundings: In **an open system,** energy can be transferred between the system and its surroundings. Organisms are open systems. They absorb light energy or chemical energy in the form of organic molecules and release heat and metabolic waste products, such as carbon dioxide, to the surroundings.

9. According to the **first law of thermodynamics**, the energy of the universe is constant. Energy can be transferred and transformed but, it cannot be created or destroyed. The first law is also known as conservation of energy. The Hydropower station does not create energy but convert potential energy stored in water into electrical energy when falling water moves the turbines.

10. The second law of thermodynamics can be stated in many ways. Let's begin with the following concept: **Every energy transfer or transformation makes the universe more disordered**. There is a quantitative measure of disorder, called **entropy**, whose value increases as disorder increases. We can now state the second law as follows: **Every energy transfer or transformation increases the entropy of the universe**. There is an unstoppable trend towards randomization. In many cases, increased entropy is evident in the physical disintegration of a system's organized structure. Consider, for example, the degradation of an unmaintained building. Much of the increasing entropy of the universe is less apparent, however, because it takes the form of an increasing amount of heat, which is the energy of random molecular motion.

11. Conversion of other forms of energy to heat **does not violate the first law of thermodynamics**. Energy has been conserved; because, heat is a form of energy, albeit in its random state. By combining the first and second laws of thermodynamics, we conclude that the quantity of energy in the universe is constant, but its quality is not. In a sense, heat is the lowest grade of energy; because, it is an

uncoordinated movement of molecules that many systems cannot harness in order to perform work. A system can only put heat to work when there is a temperature difference that results in the heat flowing from a warmer location to a cooler one. If temperature is uniform throughout a system, as it is in a living cell, then the only use for heat energy is to warm a body of matter, such as an organism.

12. Thus, an organism takes in organized forms of matter and energy -from the surroundings and replaces them with less ordered forms. For example, an animal obtains starch, proteins, and other complex molecules from the food it eats, and in turn releases carbon dioxide and water, relatively small and simple molecu1es that store less chemical energy than the food. The depletion of chemical energy is accounted by heat generated during metabolism.

13. Cells create ordered structures from less organized starting materials. For example, amino acids .are ordered into the specific sequences of polypeptide chains. Complex organisms evolve from simpler ancestors. But, this high degree of organization in no way violates the second law; because, the entropy of a particular system, such as an organism, may actually decrease, as long as the total entropy of the universe -the system plus its surroundings - increases. Thus, organisms are islands of low entropy in an increasingly random universe. The evolution of biological order is perfectly consistent with the laws of thermodynamics.

14. Bioenergetics or biochemical thermodynamics deals with the study of energy changes (transfer and utilization) in biochemical reactions. The reactions are broadly classified as **exergonic** (energy releasing) and **endergonic** (energy consuming). Bioenergetics is concerned with the initial and final states of energy component of the reactants and not the mechanism of chemical reactions.

15. Free energy: The energy actually available to do work (utilizable) is known as free energy. Changes in the **free energy** (ΔG) are valuable in predicting the feasibility of chemical reactions. The reactions can occur spontaneously if they are accompanied by decrease in free energy.

16. During a chemical reaction, heat may be released or absorbed. **Enthalpy (ΔH)** is a measure of the change in heat content of the reactants, compared to products.

17. Entropy (ΔS) represents a change in the randomness or disorder of reactants and products. Entropy attains a maximum as the reaction approaches equilibrium. The relation between the changes of free energy (ΔG), enthalpy (ΔH) and entropy (ΔS) is expressed as

$$\Delta G = \Delta H - T\Delta S$$

18. The term standard free energy represented by ΔG^0 (note the superscript ^{0}) is often used. It indicates the free energy change when the reactants or products are at a concentration of 1 mol/l at pH 7.0.

19. Negative and positive ΔG: If free energy change (ΔG) is represented by a negative sign, there is a loss of free energy. The reaction is said to be exergonic, and proceeds spontaneously. On the other hand, a positive ΔG indicates that energy must be supplied to the reactants. The reaction cannot proceed spontaneously and is endergonic in character.

The hydrolysis of ATP is a classical example of exergonic reaction

$$ATP + H_2O \rightarrow ADP + Pi \ (\Delta G^o = -7.3 \text{ Cal/mol})$$

The reversal of the reaction (ADP + Pi → ATP) is endergonic and occurs only when there is a supply of energy of at least 7.3 Cal/mol (ΔGo is positive).

20. The **free energy change becomes zero** ($\Delta G = 0$) when a reaction is at **equilibrium.** At a constant temperature and pressure, ΔG is dependent on the actual concentration of reactants and products. For the conversion of reactant A to product B (A→ B), the following mathematical relation can be derived

$$\Delta G = \Delta G^0 + RT \ln \frac{[B]}{[A]}$$

where ΔG^0 = Standard free energy change
R = Gas constant (1.987 Cal/mol)
T = Absolute temperature (273 + °C)
ln = Natural logarithm
[B] = Concentration of product
[AI = Concentration of reactant.

21. Energy producing and energy utilizing systems

- **Catabolic pathways** - generation of energy by transformation (glycolysis) or oxidation (oxidative phosphorylation) of ingested or stored fuels.
- **Anabolic pathways** - utilization of energy for biosynthetic purposes (DNA synthesis, protein synthesis, etc.)
- **ATP** - the link between the two pathways. **Phosphoanhydride bonds (~P)** as high energy bonds(~7 kcal/mol). Other nucleotides which are used as energy currency
  a. GTP: gluconeogenesis and protein synthesis
  b. CTP: lipid synthesis
  c. UTP: glycogen synthesis
  These all depend on ATP formation. Enzyme nucleoside diphosphate kinase utilize ATP to convert GTP, CTP or UTP.
- Some other **Energy rich bonds** having a high free energy of hydrolysis (**> 7 kcal/mole**)
  a. 1,3 bis-phosphoglycerate (mixed acid anhydrides)
  b. Phosphoenolpyruvate (enol phosphates)
  c. Creatine phosphate (phosphoguanidines
  d. Pyrophosphate (phosphoric acid anhydrides)
  e. Acetyl CoA (thiol esters)

D4 Colligative properties

Properties of solutions that depend on the number of molecules present and not on the kind of molecules are called colligative properties. These properties include boiling point elevation, freezing point depression, and osmotic pressure. Historically, colligative properties have been one means for determining the molecular weight of unknown compounds. In this chapter we discuss using colligative properties to measure the molecular weight of polymers. Because colligative properties depend on the number of molecules, we expect, and will show, that colligative property experiments give a number average molecular weight.

Relative lowering of vapour pressure
As seen earlier,

$$\frac{P_A^0 - P_A}{P_A^0} = \chi_B \quad \text{or} \quad \frac{P_A^0 - P_A}{P_A^0} = \frac{n_B}{n_A + n_B}$$

$$= \frac{\frac{w_B}{M_B}}{\frac{w_A}{M_A} + \frac{w_B}{M_B}}$$

For dilute solution, n_B is negligible in comparison with n_A

$$\therefore \frac{P_A^0 - P_A}{P_A^0} = \frac{n_B}{n_A}$$

$$\frac{P_A^0 - P_A}{P_A^0} = \frac{w_B}{M_B} \times \frac{M_A}{w_A}$$

or

$$\frac{\Delta P_A}{P_A^0} = \frac{w_B}{M_B} \times \frac{M_A}{w_A}$$

By experimentally measuring relative lowering of vapour pressure in a solution of known mass of solute and molar mass of solvent and its weight, molar mass of solute can be determined.

$$\frac{P_A^0 - P_A}{P_A^0} = \frac{w_B}{M_B} \times \frac{M_A}{w_A} = \frac{w_B}{M_B} \times \frac{1}{w_A} \times M_A = n_B \times \frac{1}{w_A} \times M_A$$

If w_A is mass of solvent in kilogram, then $n_B \times \frac{1}{w_A}$ is molality of solution and M_A is the molar mass of the solvent in kg mol^{-1}.

$$\therefore \frac{P_A^0 - P_A}{P_A^0} = m \times M_A$$

$$\frac{P_A^0 - P_A}{P_A^0} \propto m$$

i.e.

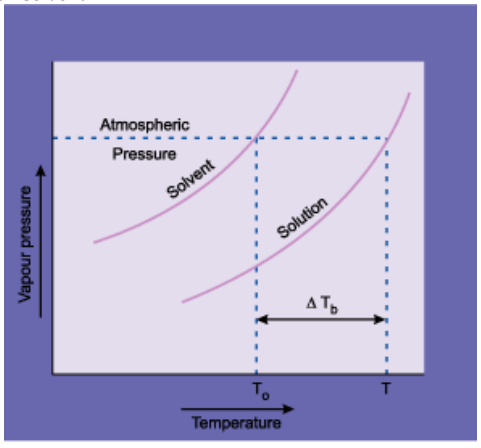
Hence, we can say that relative lowering of vapour pressure is a colligative property as colligative properties are proportional to molality.

Elevation in boiling point

Before understanding the elevation in boiling point, we should know what is boiling point.

Boiling point of a liquid can be defined as that temperature at which its vapour pressure becomes equal to that of the atmospheric pressure.

A solution of non-volatile solute has higher boiling point than solvent.



Thus, elevation in boiling point, $\Delta T_b = T - T_o$

As solution containing non-volatile solute has lower vapour pressure than solvent, the boiling point of solution is higher than solvent as evident from the above figure.

Being a colligative property, ΔT_b is proportional to molality.

i.e. $\Delta T_b \propto m$

$\Delta T_b = K_b \times m$

Where K_b is constant.

If m = 1, then $\Delta T_b = K_b$

Thus,

K_b is defined as elevation in boiling point of 1 molal solution. It is called molal elevation constant or ebullioscopic constant.

Units of K_b

$$K_b = \frac{\Delta T_b}{m} = \frac{K}{mol\,kg^{-1}} = K\ kg\ mol^{-1}$$

For numericals, the expression for elevation in boiling point is given as,

$$\Delta T_b = K_b \times \frac{w_B}{M_B} \times \frac{1}{w_A} \times 1000$$

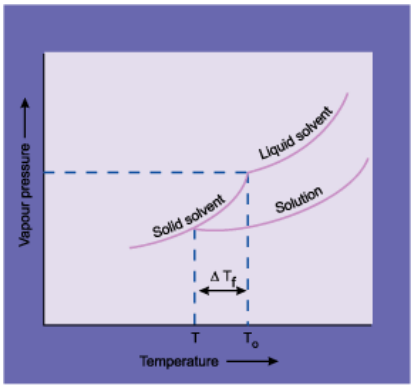
K_b for water is 0.52 K kg mol^{-1} means 1 mole of a substance (solute) in 1 kg of water increases its boiling point by 0.52 K. Thus, boiling point of 1 molal solution in water = 373 K + 0.52 K = 373.52 K.

An aqueous solution of ethylene glycol is used as coolant in cars in hot regions. It raises the boiling point of water and decreases it vapour pressure. Thus, water evaporates slowly and lasts longer in the radiator.

Depression in freezing point

Freezing point of a substance is the temperature at which the vapour pressure of its solid form is same as the vapour pressure of its liquid form or in other words, the temperature at which its solid form is in equilibrium with its liquid form.

When non-volatile solute is added to pure solvent, the freezing point of solvent lowers.



Thus, depression in freezing point, $\Delta T_f = T_o - T$

Being a colligative property, ΔT_f is proportional to molality.

i.e. $\Delta_f \propto m$

$\Delta T_f = K_f \times m$

Where K_f is constant.

If m =1, then $\Delta T_f = K_f$

K_f is defined as depression in freezing point of 1 molal solution. It is called molal depression constant or cryoscopic constant.

Units of K_f

$$K_f = \frac{\Delta T_f}{m} = \frac{K}{mol\ kg^{-1}} = K\ kg\ mol^{-1}$$

For numericals, the expression for depression in freezing point is given as,

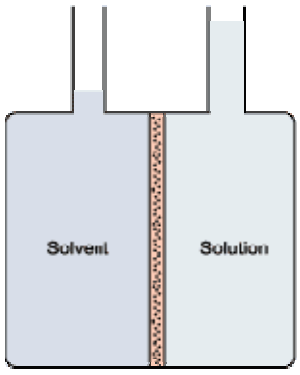
$$\Delta T_f = K_f \times \frac{w_B}{M_B} \times \frac{1}{w_A} \times 1000$$

K_f for water is 1.86 K kg mol^{-1} means 1 mole of solute in 1 kg water depresses it freezing point by 1.86 K. Thus, freezing point of 1 molal solution in water = 273 – 1.86 = 271.14 K.

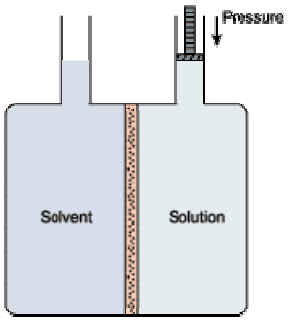
An aqueous solution of ethylene glycol is used as antifreeze in car engines in cold regions. Freezing point of this solution is lower than the prevalent temperature. Thus, the water in car radiator does not freeze. Mixtures of ice and salts are used for producing low temperature for various purposes.

Osmotic pressure

When dried fruits, beans and vegetables are kept in water, they slowly swell. This is because water enters through the skin of fruits and vegetables and this process is called osmosis. Here, skin of fruits and vegetables act as semi-permeable membrane. Other examples of semi-permeable membrane are egg membrane, cellulose acetate membrane, etc.



Flow of solvent molecules through semi-permeable membrane towards solution is osmosis.



To stop the process of osmosis, extra pressure is applied from the solution side. It is **osmotic pressure** (Π). The units of osmotic pressure are atm or mm Hg or Nm^{-2} or kPa.

Π increases with concentration of solute, i.e. $\frac{n_B}{V}$.

i.e. $\Pi \propto \frac{n_B}{V}$

Π increases with temperature T.

i.e. $\Pi \propto T$

Therefore, $\Pi \propto \frac{n_B}{V} \times T$

or $\Pi = R\frac{n_B}{V}T$

Where R is constant called solutions constant (same as ideal gas constant).

$$\Pi = \frac{n_B}{V} \times R \times T$$

$$\Pi = C \times R \times T$$

where C is the number of moles of solute present in one litre solution.

This method is used for the determination of molar masses of polymers, proteins and other macromolecules for which other methods (i.e. elevation in boiling point and depression in freezing point) cannot be used as the changes produced in boiling point and freezing point are too small to be measured accurately, whereas the osmotic pressure generated is large enough for accurate measurement.

E1. BIOENERGETICS (CONCEPT OF FREE ENERGY)

Energy: Some basic principles:

The two concepts most basic to science are matter and energy. We shall introduce these concepts, by defining matter as "**anything that has mass and takes up space**". Energy, however, is more abstract. It can only be described and measured by how it affects matter. Recall that energy is the capacity to do work - that is, to move matter against opposing forces, such -as gravity and friction. Put another way, energy is the ability to change the way a collection of matter is arranged.

Forms of Energy:

Anything that moves possesses a form of energy called kinetic energy. Any moving object, including an atom or molecule, has **kinetic energy**- the energy of motion. Moving objects perform work by imparting motion to other matter: A pool player uses the motion of the cue stick to push the cue ball, which in turn moves the other balls; water gushing through a dam turns turbines; electrons flowing along a wire run household appliances; the contraction of leg muscles pushes bicycle pedals. **Heat**, or **thermal energy**, is kinetic

energy that results from the random movement of molecules. **Light** also represents energy, which can be harnessed to perform work, such as powering photosynthesis in green plants.

A resting object, not presently at work, may also possess energy, which is the capacity to do work. Stored energy, or **potential energy**, is energy that matter possesses because of its location or arrangement. Water behind a dam, for instance, stores energy because of its altitude. Chemical energy, a form of potential energy especially important to biologists, is stored in molecules because of the arrangement of the atoms that are bonded together.

Energy Transformations

Energy can be converted from one form to another. Chemical energy can be tapped when chemical reactions rearrange the atoms of molecules in such a way that potential energy stored in the molecules is converted to kinetic energy. This transformation occurs, for example, in the engine of an automobile when the hydrocarbons of gasoline react explosively with oxygen releasing the energy that pushes the pistons. Similarly, chemical energy fuels organisms. Cellular respiration and other anabolic pathway unleash energy stored in sugar and other complex molecules and make that energy available for cellular work. The chemical energy stored in the fuel molecules had itself been converted from light energy by plants during photosynthesis.

The Free-Energy Concept: A Criterion for Spontaneous Change

How can we predict what can and cannot occur in nature? How can we distinguish the possible from the impossible? We know from experience that certain events occur spontaneously and others do not. For instance, we know that water flows downhill, that objects of opposite charge move toward each other, that an ice cube melts at room temperature, and that a sugar cube dissolves in water. But, explaining why these processes occur spontaneously is tricky.

Let's begin by defining a spontaneous process as a change that can occur without outside help. A spontaneous change can be harnessed in order to perform work. The down flow of water can be used to turn a turbine in a power plant, for example. A process that cannot occur on its own is said to be non-spontaneous; it will happen only if an external energy source is added. Water moves uphill only when a windmill, or some other machine, pumps the water against gravity, and a cell must expend energy to synthesize a protein from amino acids.

When a spontaneous process occurs in a system, the stability of that system increases. Unstable systems tend to change in such a way that they become more stable. A body of elevated water, such as a reservoir, is less stable than the same water at sea level. A system of charged particles is less stable when opposite charges are apart than when they are together. But, in situations less familiar to us, how can we predict which changes lead to greater stability in a system? Which changes are spontaneous? The **concept of entropy** teaches us that a process can only occur spontaneously if it increases the disorder of the universe. This principle is helpful in theory; but, it does not give us a practical criterion to apply to biological systems because, it requires that we measure changes in the surroundings. We need some standard for spontaneity that is based on the system alone. That criterion is called **free energy**. The concept of free energy is not easy to grasp; but the effort is worthwhile; because, the idea can be applied to many biological problems.

Free energy is the portion of a system's energy that can perform work when temperature is uniform throughout the system, as in a living cell. A system's quantity of free energy is symbolized by the letter "**G**". There are two components to "**G**": the system's total energy (symbolized by **H**) and its entropy (symbolized by **S**). Free energy is related to these factors in the following way:

$$G=H-TS$$

T stands for absolute temperature (°C + 273). Notice that temperature amplifies the entropy term of the equation. This makes sense, because, as we learned earlier, temperature measures the intensity of random molecular motion (heat), which tends to disrupt order. What does this equation tell us about free energy? Not all the energy stored in a system is availab1e for work. The systems disorder, the entropy factor decreases the availability of energy to perform any work.

How does the concept of free energy help us determine whether a particular process can occur spontaneously? Think of free energy, **G** in the above equation, as a measure of a system's instability - its tendency to change to a more stable state. Systems that are rich in energy, such as stretched springs or separated charges, are unstable. So are highly ordered systems, such as complex molecules. Thus, those systems that tend to change spontaneously to a more stable state are those that have high energy, low entropy, or both. The free-energy equation weighs these two factors, which are consolidated in the system's G content. Now, we can state a versatile criterion for spontaneous change:

In any spontaneous process, the free energy of a system decreases.

The change in free energy as a system goes from a starting state to a different state is represented by ΔG:

$$\Delta G = G_{\text{final state}} - G_{\text{Starting state}}$$

or put another way:

$$\Delta G=\Delta H -T\Delta S$$

T stands for absolute temperature (in Kelvin units, **K**, °C + 273). For a process to occur spontaneously, the system must either give up energy (a decrease in **H**), give up order (an increase in **S**), or both. When these changes in H and S are tallied, ΔG must have a negative value. The greater this decreases in free energy, the greater the maximum amount of work the spontaneous process can perform. This is a formal, mathematical way of stating the obvious: Nature runs "downhill" (downhill in the broad metaphorical sense of a loss in useful energy - the capacity to perform work).

Two Laws of Thermodynamics

The study of the energy transformations that occur in a collection of matter is called **thermodynamics**. Scientists use the word system to denote the collection of matter under study and refer to the rest of the universe - everything outside the system - as the surroundings. **A closed system**, such as that approximated by liquid in a thermos bottle, is isolated from its surroundings: In **an open system,** energy can be transferred between the system and its surroundings. Organisms are open systems. They absorb light energy or chemical energy in the form of organic molecules and release heat and metabolic waste products, such as carbon dioxide, to the surroundings.

According to the **first law of thermodynamics**, the energy of the universe is constant. Energy can be transferred and transformed but, it cannot be created or destroyed. The first law is also known as conservation of energy. The Hydropower station does not create energy but convert potential energy stored in water into electrical energy when falling water moves the turbines.

The second law of thermodynamics can be stated in many ways. Let's begin with the following concept: **Every energy transfer or transformation makes the universe more disordered**. There is a quantitative measure of disorder, called **entropy**, whose value increases as disorder increases. We can now state the second law as follows: **Every energy transfer or transformation increases the entropy of the universe**. There is an unstoppable trend towards randomization. In many cases, increased entropy is evident in the physical disintegration of a system's organized structure. Consider, for example, the degradation of an unmaintained building. Much of the increasing entropy of the universe is less apparent, however, because it takes the form of an increasing amount of heat, which is the energy of random molecular motion.

In most energy transformations, ordered forms of energy are at least partly converted to heat. Only about 25% of the chemical energy stored in the fuel tank of an automobile is transformed into the motion of the car, the remaining 75% is lost from the car engine as heat, which dissipates rapidly through the surroundings. In performing various kinds of work, living cells unavoidably convert organized forms of energy to heat (This makes a room crowded with people uncomfortably warm).

In machines and organisms, even energy that performs useful work is eventually converted to heat. The organized energy of an automobile's forward movement becomes heat when the friction of the brakes and tyres stops the car. Conversion to heat is the fate of all the chemical energy.

Conversion of other forms of energy to heat **does not violate the first law of thermodynamics**. Energy has been conserved; because, heat is a form of energy, albeit in its random state. By combining the first and second laws of thermodynamics, we conclude that the quantity of energy in the universe is constant, but its quality is not. In a sense, heat is the lowest grade of energy; because, it is an uncoordinated movement of molecules that many systems cannot harness in order to perform work. A system can only put heat to work when there is a temperature difference that results in the heat flowing from a warmer location to a cooler one. If temperature is uniform throughout a system, as it is in a living cell, then the only use for heat energy is to warm a body of matter, such as an organism.

Thus, an organism takes in organized forms of matter and energy -from the surroundings and replaces them with less ordered forms. For example, an animal obtains starch, proteins, and other complex molecules from the food it eats, and in turn releases carbon dioxide and water, relatively small and simple molecu1es that store less chemical energy than the food. The depletion of chemical energy is accounted by heat generated during metabolism.

Cells create ordered structures from less organized starting materials. For example, amino acids .are ordered into the specific sequences of polypeptide chains. Complex organisms evolve from simpler ancestors. But, this high degree of organization in no way violates the second law; because, the entropy of a particular system, such as an organism, may actually decrease, as long as the total entropy of the universe -the system plus its surroundings - increases. Thus, organisms are islands of low entropy in an increasingly random universe. The evolution of biological order is perfectly consistent with the laws of thermodynamics.

E2. THERMODYNAMIC PRINCIPLES IN BIOLOGY

Free energy change: If we consider a system consisting of two compounds A and B which react to give C:

$$A+B \rightarrow C$$

By reacting, the system can exchange work or energy (heat, light) with the immediate surroundings. For example, it can receive energy and produce an equivalent work,

At the time of mixing, A+ B represents the initial state, and C, the product formed, represents the final state.

Each of these two states, is characterized by an internal energy level (U_1 and U_2, respectively) which depends en the chemical nature of the constituents (with their atoms, their bonds), their concentrations and their position in space (absolute temperature T, pressure P, Volume V). **The internal energy change**, denoted by Δ**U**, resulting from a change of molecular structures (new bonds and different distribution of electrons), depends only on the initial state and final state: $\Delta U = U_2 - U_1$.

It is the maximum quantity of energy which can be made available to accomplish a work by the transformation of A + B into C. In a constant pressure system (as in the case of many biological reactions) there can be volume changes and either a liberation or an absorption of heat.

The work done by pressure is equal to P ($V_2 - V_1$), i.e., P ΔV. The heat of reaction first defined by ΔQ is equal to the internal energy change plus the work done

$\Delta Q = \Delta U + P\Delta V$

$\Delta H = \Delta U + P\Delta V$ -- **(1)**

for a given temperature, because we know that PV = nRT (n = number of moles, R = gas constant, T = temperature, in degrees Kelvin).

ΔH is the enthalpy change of the system. **ΔH** like **ΔQ** is mostly expressed in cal or kcal/mole, but ought to be expressed in joules/mole (1 calorie = 4.18 Joules).

ΔH can be measured in a calorimeter. If **$\Delta H < 0$**, the reaction is exothermic (heat is liberated) and if **$\Delta H > 0$**, the reaction is endothermic (the system absorbs heat). In the past, it was thought that the knowledge of the sign of **Δ**H is sufficient to predict whether a reaction is spontaneous or not and it was said that an exothermic reaction is spontaneous, while an endothermic reaction is not. It was then found that, while these predictions are correct in most cases, there are exceptions, and that some spontaneous reactions are, in fact, accompanied by absorption of heat by the molecules.

One was, therefore, led to consider the entropy of the system denoted S, a function which measures the disorder of a system and the variations of which, **$\Delta S = S_2 - S_1$,** depends - like **Δ**U - on the initial state and the final state. In order to pass from a first stable state, the initial state, to a new stable state, the final state, and one must upset the structures of molecules A and B for the reaction to take place. The greater the increase in molecular disorder, the greater the entropy (for example, entropy increases while passing from a given protein to the constituent amino acids in free State). Entropy is related to the complexity of the molecules. The larger the number of atoms, the larger the number of complicated chemical groups; the larger the number of possible

structures and the greater the entropy. In a reversible transformation at constant pressure. **Δ**Q = T**Δ**S. One can then write:

ΔH = ΔG+ΔQ = ΔG+TΔS **(2)**

Or **Δ** G (free enthalpy change)= **ΔH-TΔS** **(3)**

In a reversible transformation at constant volume (ΔV = 0) the free energy change is

ΔF=ΔU-TΔS -- (4)

In biochemistry, it may admitted that in most cases **Δ**V = 0; because the volume changes for reactions taking place in solution are negligible. In equation (3) giving **Δ**G, it is, therefore, admitted that P**Δ**V= 0, so that **Δ**G becomes, like **Δ**F (equation 4), equal to **ΔU-TΔ**S and, therefore, **Δ**F becomes **Δ**G. This is particularly true if the relidions involve liquids and solids in solution (and no gas) at physiological temperatures which are low.

Three cases are possible. When **Δ**G < 0, the reaction is **exergonic** (it can be spontaneous) and free energy is available. On the contrary, if **Δ**G > 0, the reaction is **endergonic**; free energy must be supplied to the system for the reaction to take place. The reaction will take place only coupled with another exergonic reaction whose negative **Δ**G is greater in absolute value than the positive **Δ**G of the endergonic reaction. Many biochemical reactions are reversible. **Δ**G = 0 when dynamic equilibrium is reached. This is very interesting; because, for very small variations of active masses of the reaction's participants, the reaction can take place in either direction, without energy problem, as **Δ**G will vary in either direction. It may be noted that if **Δ**S is small, as T is comparatively low in biochemistry, the term T**Δ**S can be neglected and then **Δ**G = **Δ**H (see equation 2); in other word, in this case, the terms "exothermic" and "exergonic" on the one hand, and "endothermic" and "endergonic" on the other, are similar.

Application to chemical reactions:

If we consider two compounds A and B in solution at 25°C under a pressure of 1 atmosphere, each in a concentration of 1 mole/litre of solvent (1 molal) and which can react completely, giving C + D (which will, therefore, be present in the same concentration after total transformation) the free energy change is called "standard" and denoted **Δ**G_o. If **Δ**G_o< 0, the reaction is exergonic; it can take place spontaneously; but; this is not necessarily the case even if the free energy change is very large. In fact, **Δ**G_o expresses the free energy change between an initial state and a final state. However, **Δ**Go gives no information on reaction velocity. A catalyst, an enzyme for examp1e might be needed to trigger the reaction and adjust its velocity without altering its thermodynamic characteristics. Despite the highly negative value of **Δ**G_o (-686 kcal/mole) of the total oxidation reaction of g1ucose ($C_6H_{12}O_6 + 6O_2 + 6CO_2 + 6H_2O$), this substance can be left for years in the presence of the oxygen of air; without any oxidation taking place.

On the one hand while ΔG_o indicates a standard state, it does not take into consideration the special circumstances which may prevail, particularly in the biological medium and condition and activities of compounds, pH, etc. Normally reversible reaction proceeds up to the equilibrium state and ΔG takes this equilibrium state into constant but the reaction can proceed up to completion if the equilibrium state is continuously broken because the reaction products are progressively consumed by a following reaction. This is what happens in metabolic chains- here, in fact, a general state of equilibrium is observed.

On the other hand, reactions which at the limit, can be reversible *in vitro* can be reversed *in vivo*, because of metabolic conditions. Actually, the free enthalpy change ΔG or free energy change ΔF will always depend on metabolic circumstances and will be related to equilibrium states *in vivo.*

ΔGo depends on the equilibrium constant K of the reaction:
$\Delta G_o = - RT \ln K$
and that in given circumstances, the effective ΔG of a reaction is:

$$\Delta G = \Delta G_0 + RT \ln \frac{[C][D]}{[A][B]}$$

(1st factor) (2nd factor)

For a reversible reaction A +B ⇔ C + D proceeding from left to right. The ΔG of the reverse reaction will have the opposite sign.

Therefore, two factors define ΔG; on the one hand, the ΔGo which characterizes a given -reaction, and on the other hand the second factor, characteristic of given circumstances and hence related to concentration, activity, dissociation or protonation or the reaction participants, etc. When the ratio of activities of the second factor becomes equal to 1, the second factor becomes zero and $\Delta G = \Delta G_o$.

It must be noted that ΔG and ΔGo correspond to pH = 0 and ambient temperature while ΔG' and $\Delta G'_o$ represent the values in the physiological conditions (pH near 7 physiological temperature, 30 to 37^0) to be specified in each case.

Let us examine the case of a reaction whose ΔGo = -3 kcal/mole. It is known that R = 1.987 cal/mole/degree, that at 25°C, T=273+25=298^0K and that in K = 2.30 log K.

We can write :
ΔG_o= -3000 cal/mole=-1.987 x 298 x 2.30 log K, therefore
log K= -3000 / -1362 = 2.2
i.e approximately equal to 2, which corresponds to value K equal to about 100.

If on the contrary, a reaction has a ΔG_o = +3 Kcal/mole, then K will be approximately equal to 0.01, in other words, the reaction will practically not take place because, equilibrium will be reached as soon as only 1% of the reactants will have been transformed.

K	logK	ΔG (= -1392 log K) in cal/mole
0.001	- 3	+4,086
0.01	- 2	+2,724
0.1	- 1	+1,362
1	0	0
10	+1	-1,362
100	+2	-2,724
1000	+3	-4,086
10,000	+4	-5,448
100,000	+5	-6,810
1000,000	+6	-8,172

It follows that the ΔG_0 of a reaction can be calculated if one knows K, i.e., if one can measure the concentrations of the reactants and reaction products (this is possible, except when K is extremely small or extremely large; because, then, the concentrations of the reaction products or reactants arc to low to be determined accurately). Table gives the values of ΔGo for series of values of K from 10^{-3} to 10^6.

It is clear that the reactions whose equilibrium constant is > 1, are accompanied by-a decrease of free energy (this energy becomes available); these are exergonic reactions. On the contrary, if K is <1, the reaction is endergonic.

E3. GLYCOLYSIS

Glycolysis is derived from the Creek words (glycose--sweet or sugar; lysis-dissolution). It is a universal pathway in the living cells. This pathway is often referred to as Embden-Meyerhof pathway (E.M. pathway) in honour of the two biochemists who made a major contribution to the knowledge of glycolysis.

Glycolysis is defined as "*the sequence of reactions converting glucose (or glycogen) to pyruvate or lactate, with the production of ATP*".

Salient features:

1. Glycolysis takes place in all cells of the body. The enzymes of this pathway are present in the cytosomal fraction of the cell.
2. Glycolysis occurs in the absence of oxygen (anaerobic) or in the presence of oxygen (aerobic). Lactate is the end product under anaerobic condition. In the aerobic condition, pyruvate is formed, which is then oxidized to CO_2 and H_2O.
3. Generally, Embden-Meyerhof pathway is an emergency energy-yielding pathway for cells in the absence of oxygen.
4. The occurrence of glycolysis is a prerequisite for the aerobic oxidation of carbohydrates, the latter takes place in the cells possessing mitochondria.
5. Glycolysis is a major pathway for ATP synthesis in tissues lacking mitochondria, e.g. erythrocytes, cornea, lens etc.
6. In some other tissues which have relatively few mitochondria (e.g. testes, leucocytes and kidney medulla), glycolysis is significant for ATP production.
7. Glycolysis is very essential for brain which is dependent on glucose for energy. The glucose in brain has to undergo glycolysis before it is oxidized to CO_2 and H_2O.
8. Glycolysis (anaerobic) may be summarized by the net reaction

   Glucose + 2ADP + 2Pi → 2Lactate + 2ATP
9. The intermediates formed in glycolysis are useful for the synthesis of non-essential amino acids and glycerol, the latter used for fat formation.
10. Reversal of glycolysis along with the alternate arrangements made at the irreversible steps will result in the synthesis of glucose (gluconeogenesis).

Reactions of glycolysis

The sequence of reactions of glycolysis are given in **Figure 1**. The pathway can be divided into three distinct phases.
A. Energy investment phase or priming stage
B. Splitting phase
C. Energy generation phase.

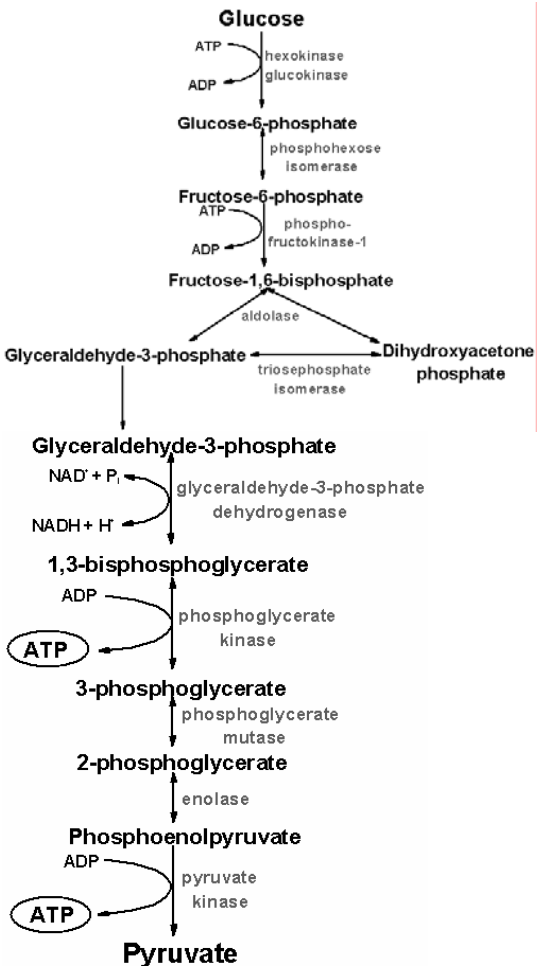
A. Energy investment phase

1. Glucose is phosphorylated to glucose 6 phosphate by hexokinase or glucokinase. This is an irreversible reaction, dependent on ATP and Mg^{2+}. The enzyme hexokinase is present in almost all the tissues. It catalyses the phosphorylation of various hexoses (fructose, man nose etc.), has low Km for substrates (about 0.1 mM) and is inhibited by glucose 6-phosphate.

Glucokinase present in liver, catalyses the phosphorylation of only glucose, has high Km for glucose (10 mM) and is not inhibited by glucose 6-phosphate.

Due to high affinity (low K_m), glucose is utilized by hexokinase even at low concentration, whereas glucokinase acts only at higher levels of glucose i.e., after a meal when blood glucose concentration is above 100 mg/dl.

Glucose 6-phosphate is impermeable to the cell membrane. It is a **central molecule with a variety of metabolic fates**-glycolysis, glycogenesis, gluconeogenesis and pentose phosphate pathway.
2. Glucose 6-phosphate undergoes isomerization to give fructose 6-phosphate in the presence of the enzyme phosphohexose isomerase and Mg^{2+}.
3. Fructose 6-phosphate is phosphorylated to fructose 1/ 6-bisphosphate by phosphofructokinase (PFK). This is an irreversible and regulatory step in glycolysis. PFK is an allosteric enzyme, the activity of which is controlled by several allosteric modulators.



B. Splitting phase

4. The six carbon fructose 1, 6-bisphosphate is split (hence the name glycolysis) to two three-carbon compounds, glyceraldehyde 3phosphate and dihydroxyacetone phosphate by the enzyme aldolase (fructose 1/6-bisphosphate aldolase).

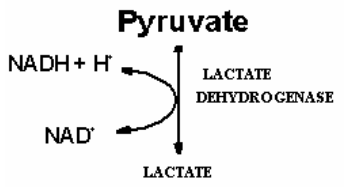
5. The enzyme phosphotriose isomerase catalyses the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. Thus, two molecules of glyceraldehyde 3-phosphate are ob~ained from one molecule of glucose. This isomerase enzyme is inhibited by **bromo-hydroxyl-acetone phosphate**.

C. Energy generation phase

6. Glyceraldehyde 3-phosphate dehydrogenase converts glyceraldehyde 3-phosphate to 1, 3-bisphosphoglycerate. This enzyme brings about two reactions, namely oxidation, where the hydrogens are transferred to NAD+ and phosphorylation of substrate adding one phosphate moiety. This step is important as it is involved in the formation of NADH + H+ and a high energy compound 1, 3-bisphosphoglycerate. **lodoacetate** and **arsenate** inhibit the enzyme glyceraldehyde 3-phosphate dehydrogenase. In aerobic condition/ NADH passes through the electron transport chain and 6 ATP (2 x 3 ATP) are synthesized by oxidative phosphorylation.
7. The enzyme phosphoglycerate kinase acts on 1, 3-bisphosphoglycerate resulting in the synthesis of ATP and formation of 3-phosphoglycerate. This step is a good example of **substrate level phosphorylation**, since ATP is synthesized from the substrate without the involvement of electron transport chain. **Phosphoglycerate kinase reaction is reversible, a rare example among the kinase reactions**.
8. 3-Phosphoglycerate is converted to 2-phosphoglycerate by phosphoglycerate mutase. This is an isomerization reaction, involving the transfer of phosphate group from position 3 to position 2 in the substrate. The enzyme requires Mg^{2+}.
9. The high energy compound pnosphoenol pyruvate is generated from 2-phosphoglycerate by the enzyme enolase. This enzyme requires Mg^{2+} or Mn^{2+} and is inhibited by fluoride. For blood glucose estimation in the laboratory, **fluoride** is added to the blood to prevent glycolysis by the cells, so that blood glucose is correctly estimated.
10. The enzyme pyruvate kinase catalyses the transfer of high energy phosphate from phosphoenol pyruvate to ADP, leading to the formation of ATP. This step also is a **substrate level phosphorylation**. Pyruvate kinase requires K^+ and either Mg^{2+} or Mn^{2+}. This reaction is irreversible.

Conversion of pyruvate to lactate-significance

The fate of pyruvate produced in glycolysis depends on the presence or absence of oxygen in the cells. Under anaerobic conditions (lack of O_2) pyruvate is reduced by NADH to lactate in presence of the enzyme lactate dehydrogenase (competitive inhibitor-oxamate). The NADH utilized in this step is obtained from the reaction catalysed by glyceraldehyde 3-phosphate dehydrogenase. The formation of lactate allows the regeneration of NAD+ which can be reused by glyceraldehyde 3-phosphate dehydrogenase so that glycolysis proceeds even in the absence of oxygen to supply ATP.



The occurrence of uninterrupted glycolysis is very essential in skeletal muscle during laborious exercise where oxygen supply is very limited. **Glycolysis in the erythrocytes leads to lactate production**, since mitochondria-the centres for aerobic oxidation-are absent. Brain, retina, skin, renal medulla and gastrointestinal tract derive most of their energy from glycolysis.

Production of ATP in glycolysis

Under anaerobic conditions, **2 ATP** are synthesized while, under aerobic conditions, **8 or 6 ATP** are synthesized-depending on the shuttle pathway that operates.

Glycolysis and shuttle pathways

Glycolysis takes place in the cytosol. The NADH produced by the enzyme, glyceraldehyde 3-phosphate dehydrogenase, is utilized later by the enzyme lactate dehydrogenase and NAD^+ is regenerated. So far as the synthesis and utilization of reducing equivalents are coupled, glycolysis proceeds uninterrupted and the end product is lactate. The perfect coupling is possible only in cells that lack mitochondria or under anaerobic conditions. However, in the presence of mitochondria and oxygen, the NADH produced in glycolysis can participate in the shuttle pathways for the synthesis of ATP.

If the cytosolic NADH uses **malate-aspartate shuttle**, 3 ATP are generated from each molecule of NADH. This is in contrast to **glycerol-phosphate shuttle** that produces only 2 ATP.

Irreversible steps in glycolysis

Most of the reactions of glycolysis are reversible. However, the three steps catalysed by the enzymes **hexokinase (or glucokinase), phosphofructokinase** and **pyruvate kinase**, are irreversible. These three stages mainly regulate glycolysis. The reversal of glycolysis, with alternate arrangements made at the three irreversible stages, leads to the synthesis of glucose from pyruvate (gluconeogenesis).

Regulation of glycolysis

The three enzymes namely hexokinase (and glucokinase), phosphofructokinase and pyruvate kinase, catalyzing the irreversible reactions regulate glycolysis.

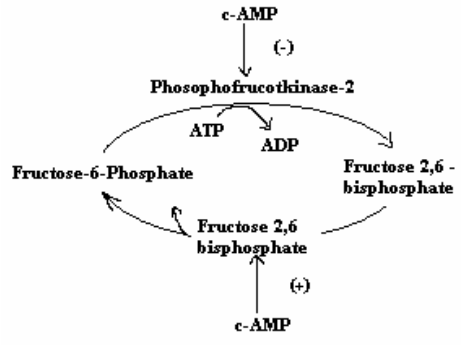
Hexokinase is inhibited by glucose-6-phosphate. This enzyme prevents the accumulation of glucose-6-phosphate due to product inhibition Glucokinase, which specifically phosphorylates glucose, probably through the involvement insulin, induces glucokinase.

Phosphofructokinase (PFK) is the most important regulatory enzyme in glycolysis. This enzyme catalyses the **rate limiting committed step**. PFK is an allosteric enzyme regulated by allosteric effectors. ATP, citrate and H^+ ions (low pH) are the most important allosteric inhibitors, whereas, fructose 2, 6-bisphosphate, AMP and Pi are the allosteric activators.

Role of fructose 2, 6-bisphosphate

Fructose 2, 6-bisphosphate (F2, 6BP) is considered to be the most important regulatory factor (activator) for controlling PFK and, ultimately, glycolysis in the liver. F2, 6BP is synthesized from fructose 6-phosphate by the enzyme phosphofructokinase called PFK-2 (PFK-l is the glycolytic enzyme). F2, 6BP is hydrolysed by fructose 2, 6-bisphosphatase. The function of synthesis and degradation of F2, 6BP is brought out by a single enzyme (same polypeptide with two active sites) which is referred to as **bifunctional enzyme**. In fact, the combined name of phosphofructokinase-2/ fructose 2, 6-bisphosphatase is used to refer to the enzyme that synthesizes and degrades F2, 6BP.

The activity of PFK-2 and fructose 2,6-bisphosphatase is controlled by covalent modification which, in turn, is regulated by cyclic AMP (cAMP is the second messenger for certain of hormones). Cyclic AMP brings about dephospho-rylation of the bifunctional enzyme, resulting in inactivation of active site responsible for the synthesis of F2, 6BP but activation of the active site responsible for the hydrolysis of F2, 6BP.



Regulation of fructose 2,6-bisphosphate

Pyruvate kinase also regulates glycolysis. This enzyme is inhibited by ATP and activated by Fl, 6BP. Pyruvate kinase is active (a) in dephosphorylated state and inactive (b) in phosphorylated state. Inactivation of pyruvate kinase by phosphorylation is brought about by cAMP-dependent protein kinase. The hormone-glucagon inhibits hepatic glycolysis by this mechanism.

Pasteur Effect:

The **inhibition of glycolysis by oxygen** (aerobic condition) is known as Pasteur effect. This effect was discovered by Louis Pasteur, more than a century ago, while studying fermentation by yeast. He observed that when anaerobic yeast cultures (metabolizing yeast) were exposed to air, the utiliziatiol1 of glucose decreased by nearly sevenfold.

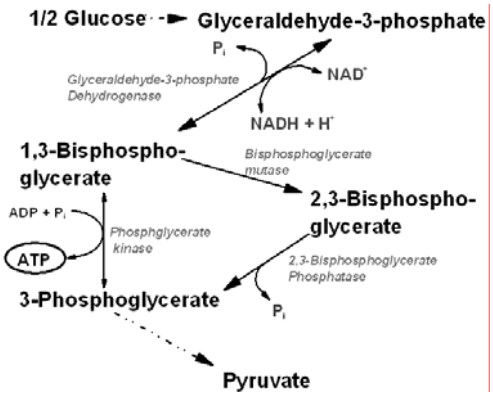
In the aerobic condition, the levels of glycolytic intermediates from fructose 1, 6-bisphosphate onwards decrease while the earlier intermediates accumulate. This clearly indicates that Pasteur effect is due to the inhibition of the enzyme phosphofructokinase. The inhibitory effect of citrate and ATP (produced in the presence of oxygen) on phosphofructokinase explains the Pasteur effect.

Crabtree effect:

The phenomenon of **inhibition of oxygen consumption by the addition of glucose** to tissues having high aerobic glycolysis is known as Crabtree effect. Basically, this is opposite to that of Pasteur Effect. Crabtree effect is due to increased competition of glycolysis for inorganic phosphate (Pi) and NAD^+ which limits their availability for phosphorylation and oxidation.

RAPAPORT-LEUBERING CYCLE

This is a supplementary pathway to glycolysis which is operative in the erythrocytes of man and other mammals. Rapaport-Leubering cycle is mainly concerned with the synthesis of 2, 3bisphosphoglycerate (2, 3-BPG) in the RBC 1, 3Bisphosphoglycerate (1, 3-BPG) produced in glycolysis is converted to 2, 3-BPG by the enzyme 2,3 - bisphosphoglycerate mutase (Fig. 13.5). 2, 3BPG is hydrolysed to 3-phosphoglycerate by bisphosphoglycerate phosphatase. [Note: There is a difference between the usages-bisphosphate and diphosphate. A bisphosphate has two phosphates held separately (e.g. 2, 3-BPG), in contrast to diphosphate (e.g. ADP) where the phosphates are linked together].



It is now believed that bisphosphoglycerate mutase is a bifunctional enzyme with mutase and phosphatase activities catalysed by two different sites present on the same enzyme.

About 15-25% of the glucose converted to lactate in erythrocytes goes via 2, 3-BPG synthesis.

Significance of 2, 3-BPG

1. Production of 2, 3-BPG allows the glycolysis to proceed without the synthesis of ATP. This is advantageous to erythrocytes since glycolysis occurs when the need for ATP is minimal. RapaportLeubering cycle is, therefore, regarded as a shunt pathway of glycolysis to dissipate or waste the energy not needed by erythrocytes.

2. 2, 3- BPG, however, is not a waste molecule in RBC It combines with hemoglobin (Hb) and reduces Hb affinity with oxygen. Therefore, in the presence of 2, 3-BPG, oxyhemoglobin unloads more oxygen to the tissues. Increase in erythrocytes 2, 3-BPG is observed in hypoxic condition, high altitude, fetal tissues, anemic conditions etc. In all these cases, 2, 3-BPG will enhance the supply of oxygen to the tissues.

3. Glycosis in the erythrocytes is linked with 2, 3-BPG production and oxygen transport. In the deficiency of the enzyme hexokinase, glucose is not phosphorylated, hence the synthesis and concentration of 2, 3-BPG are low in RBC The hemoglobin exhibits high oxygen affinity in hexokinase-defective patients. On the other hand, in the patients with pyruvate kinase deficiency, the level of 2, 3-BPG in erythrocytes is high, resulting in low oxygen affinity.

CONVERSION OF PYRUVATE TO ACETYL CoA

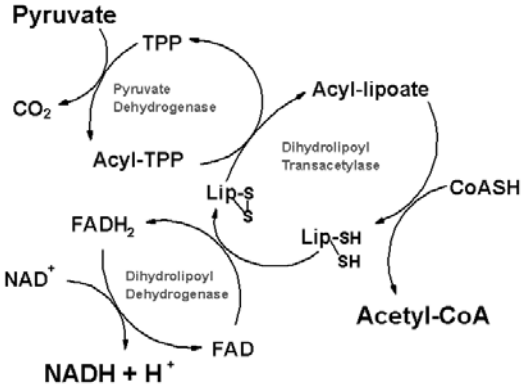
Pyruvate is converted to acetyl CoA by oxidative decarboxylation. This is an irreversible reaction, catalysed by a multienzyme complex, known as pyruvate dehydrogenase complex (PDH), which is found only in the mitochondria. High concentrations of PDH are found in cardiac muscle and kidney. The enzyme PDH requires five cofactors (coenzymes), namely-TPP, lipoamide, FAD, coenzyme A and NAD^+ (Iipoamide contains lipoic acid linked to £-amino group of lysine). The overall reaction of PDH is

$$\text{Pyruvate} + \text{NAD+} + \text{CoA} \xrightarrow{\text{PDH}} \text{Acetyl CoA} + CO_2 + \text{NADH} + H^+$$

Reactions of PDH complex

The sequence of reactions brought about by different enzymes of PDH complex in association with the coenzymes is depicted in Fig. Pyruvate is decarboxylated to give

hydroxyethyl TPP, catalysed by PDH (decarboxylase activity).



Dihydrolipoyl transacetylase brings about the formation of acetyl lipoamide (from hydroxethylTPP) and then catalyses the transfer of acetyl group to coenzyme A to produce acetyl CoA. The cycle is complete when reduced lipoamide is converted to oxidized lipoamide by dihydrolipoyl dehydro-genase, transferring the reducing equivalents to FAD. FADH2, in turn, transfers the reduciJ!lg equivalents to NAD+ to give NADH + H+, which can pass through the respiratory chain to give 3 ATP (6 ATP from 2 moles of pyruvate formed from glucose) by oxidative phosphorylation.

The intermediates of PDH catalysed reaction are not free but bound with enzyme complex. **In mammals, the PDH complex has an approximate molecular weight of 9 x 10^6. It contains 60 molecules of dihydrolipoyltransacetylase and about 20-30 molecules each of the other two enzymes (pyruvate dehydrogenase and dihydrolipoyl dehydrogenase)**.

A comparable enzyme with PDH is aketoglutarate dehydrogenase complex of citric acid cycle which catalyses the oxidative decarboxylation of α-ketoglutarate to succinyl CoA.

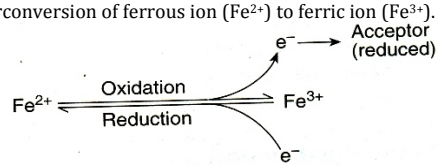
Biochemical importance of PDH

1. Lack of TPP (due to deficiency of thiamine) inhibits PDH activity resulting in the accumulation of pyruvate.
2. In the thiamine deficient alcoholics, pyruvate is rapidly converted to lactate, resulting in lactic acidosis.
3. In patients with inherited deficiency of PDH, lactic acidosis (usually after glucose load) is observed.
4. PDH activity can be inhibited by arsenic and mercuric ions. This is brought about by binding of these ions with -SH groups of lipoic acid.

[Note: Pyruvate is a key metabolite. Besides its conversion to acetyl CoA (utilized in a wide range of metabolic reactions-citric acid cycle, fatty acid synthesis etc.), pyruvate is a good substrate for guconeogenesis].

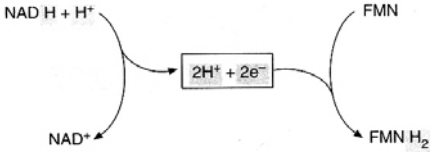
E4. BIOLOGICAL OXIDATION

Oxidation is defined as the loss of electrons and reduction as the gain of electrons. This may be illustrated by the interconversion of ferrous ion (Fe^{2+}) to ferric ion (Fe^{3+}).



The electron lost in the oxidation is accepted by an acceptor which is said to be reduced. Thus the oxidation-reduction is a tightly coupled process.

The general principle of oxidation-reduction is applicable to biological systems also. The oxidation of NADH to NAD^+ coupled with the reduction of FMN to $FMNH_2$ is illustrated.



In the above illustration, there are two redox pairs NADH/NAD+ and FMN/FMNH2. The redox pairs differ in their tendency to lose or gain electrons.

Redox potential (E_o): The oxidation-reduction potential or, simply, redox potential, is a quantitative measure of the tendency of a redox pair to lose or gain electrons. The redox pairs are assigned specific standard redox potential (E_o volts) at pH 7.0 and 25°C.

The redox potentials of some biologically important redox systems are given in Table.3. The more negative redox potential represents a greater tendency (of reductant) to lose electrons. On the other hand, a more positive redox potential indicates a greater tendency (of oxidant) to accept electrons. The electrons flow from a redox pair with more negative E_o to another redox pair with more positive E_o. The redox potential (E_o) is directly related to the change in the free energy (ΔG^0).

Table .3 Standard redox potential (E_0) of some oxidation-reduction systems

Redox pair	*E_0 Volts*
Succinate/α-ketoglutarate	–0.67
$2H^+/H_2$	–0.42
NAD^+/NADH	–0.32
$NADP^+$/NADPH	–0.32
FMN/$FMNH_2$ (enzyme bound)	–0.30
Lipoate (ox/red)	–0.29
FAD/$FADH_2$	–0.22
Pyruvate/lactate	–0.19
Fumarate/succinate	+0.03
Cytochrome b (Fe^{3+}/Fe^{2+})	+0.07
Coenzyme Q (ox/red)	+0.10
Cytochrome c_1 (Fe^{3+}/Fe^{2+})	+0.23
Cytochrome c (Fe^{3+}/Fe^{2+})	+0.25
Cytochrome a (Fe^{3+}/Fe^{2+})	+0.29
$\frac{1}{2}O_2/H_2O$	+0.82

ELECTRON TRANSPORT CHAIN

The energy-rich carbohydrates (particularly glucose), fatty acids and amino acids undergo a series of metabolic reactions and, finally, get oxidized to CO_2 and H_2O. The reducing equivalents from various metabolic intermediates are transferred to coenzymes NAD^+ and FAD to produce, respectively, NADH and $FADH_2$. The latter two reduced coenzymes pass through the electron transport chain (ETC) or respiratory chain and, finally, reduce oxygen to water.

The passage of electrons through the ETC is associated with the loss of free energy. A part of this free energy is utilized to generate ATP from ADP and Pi (Fig.3).

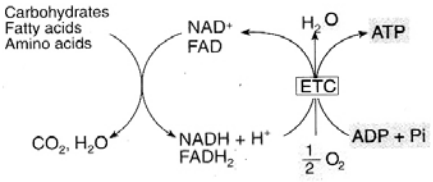


Fig. 3 : Overview of biological oxidation (ETC-Electron transport chain).

Mitochondria -the power houses of cell: The mitochondria are the centres for metabolic oxidative reactions to generate reduced coenzymes (NADH and $FADH_2$) which, in turn, are utilized in ETC to liberate energy in the form of ATP. For this reason, mitochondrion is appropriately regarded as the power house of the cell.

Mitochondrial organization: The mitochondrion consists of five distinct parts. rhese are the outer membrane, the inner membrane, the intermembrane space, the cristae and the matrix. The electron transport chain and ATP synthesizing system are located on the inner mitochondrial membrane which is a specialized structure, rich in proteins. It is impermeable to ions (H^+, K^+, Na^+) and small molecules (ADP, ATP). This membrane is highly folded to form cristae. The surface area of inner mitochondrial membrane is greatly increased due to cristae. The inner surface of the inner mitochondrial membrane possesses specialized particles (that look like lollipops), the phosphorylating subunits which are the centres for ATP production.

Mitochondrial matrix: The interior ground substance forms the matrix of mitochondria. It is rich in the enzymes responsibie for the citric acid cycle, β-oxidation of fatty acids and oxidation of amino acids.

Structural organization of respiratory chain: The inner mitochondrial membrane can be disrupted into five distinct respiratory or enzyme complexes, denoted as complex I, II, III, IV and V (Fig 6). The complexes I-IV are carriers of electrons while complex V is responsible for ATP synthesis. Besides these enzyme complexes, there are certain mobile electron carriers in the respiratory chain. These include NADH, coenzyme Q, cytochrome C and oxygen.

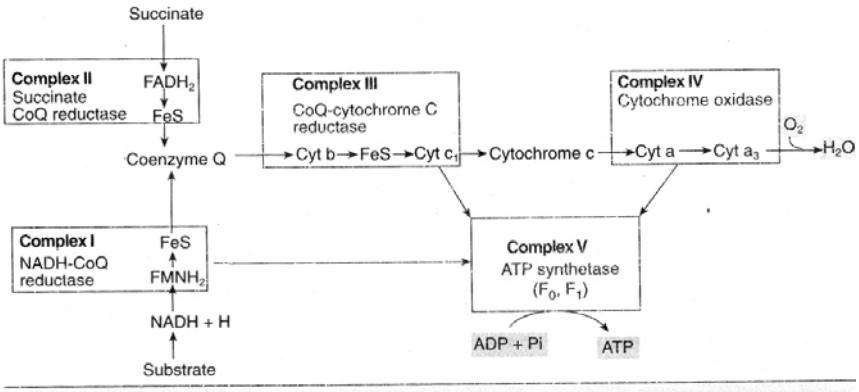


Fig. 6 : Multiprotein complexes in electron transport chain.

The enzyme complexes (I-IV) and the mobile carriers are collectively involved in the transport of electrons which, ultimately, combine with oxygen to produce water. The largest proportion of the oxygen supplied to the body is utilized by the mitochondria for the operation of electron transport chain.

Components and reactions of the electron transport chain: There are five distinct carriers that participate in the electron transport chain (ETC). These carriers are sequentially arranged (Fig. 7) and are responsible for the transfer of electrons from a given substrate to ultimately combine with proton and oxygen to form water.

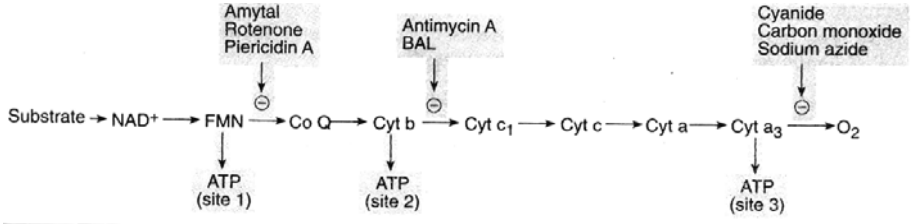


Fig. 7 : Electron transport chain with sites of ATP synthesis and inhibitors (BAL—British antilewisite).

I. Nicotinamide nucleotides: Of the two coenzymes NAD+ and NADP+ derived from the vitamin niacin, NAD+ is more actively involved in the ETC NAD^+ is reduced to **$NADH+H^+$** by dehydrogenases with the removal of two hydrogen atoms from the substrate (AH_2) The substrates include glyceraldehyde-3 phosphate, pyruvate, isocitrate, α-ketoglutarate and malate.

$$AH_2 + NAD^+ \Leftrightarrow A + NADH + H^+$$

NADPH + H^+ produced by **$NADP^+$-dependent dehydrogenase is not usually a substrate for ETC.** NADPH is more effectively utilized for anabolic reactions (e.g. fatty acid synthesis, cholesterol synthesis).

II. Flavoproteins: The enzyme NADH dehydrogenase (NADH coenzyme Q reductase) is a flavoprotein with FMN as

the prosthetic group. The coenzyme FMN accepts two electrons and a proton to form $FMNH_2$. NADH dehydrogenase is a complex enzyme closely associated with non-heme iron proteins (NHI) or iron-sulfur proteins (FeS).

$$NADH + H^+ + FMN \rightarrow 7\ NAD^+ + FMNH_2$$

Succinate dehydrogenase (succinate-coenzyme Q reductase) is an enzyme found in the inner mitochondrial membrane. It is also a flavoprotein with FAD as the coenzyme. This can accept two hydrogen atoms ($2H^+ + 2e^-$) from succinate.

$$Succinate + FAD \rightarrow Fumarate + FADH_2$$

III. Iron-sulfur proteins: The iron-sulfur (FeS) proteins exist in the oxidized (Fe^{3+}) or reduced (Fe^{2+}) state. About half a dozen FeS proteins connected with respiratory chain have been identified. However, the mechanism of action of iron-sulfur proteins in the ETC is not clearly understood. One FeS participates in the transfer of electrons from FMN to coenzyme Q. Other FeS proteins associated with cytochrome b and cytochrome c participate in the transport of electrons.

IV. Coenzyme Q: Coenzyme Q is also known as ubiquinone since it is ubiquitous in living system. It is a quinone derivative with a variable isoprenoid side chain. The mammalian tissues possess a quinone with 10 isoprenoid units which is known as coenzyme Q_{10} (CoQ_{10}). Coenzyme Q is a lipophilic electron carrier. It can accept electrons from FMNH2 produced in the ETC by NADH dehydrogenase or $FADH_2$ produced outside ETC (e.g. succinate dehydrogenase, acyl CoA dehydrogenase). **Coenzyme Q is not found in mycobacteria**. **Vitamin K performs similar function as coenzyme Q in these organisms**. Coenzyme Q has no known vitamin precursor in animals. It is directly synthesized in the body.

V. Cytochromes: The cytochromes are conjugated proteins containing heme group. The latter consists of a porphyrin ring with iron atom. The heme group of cytochromes differ from that found in the structure of hemoglobin and myoglobin. The iron of heme in cytochromes is alternately oxidized (Fe^{3+}) and reduced (Fe^{2+}), which is essential for the transport of electrons in the ETC. This is in contrast to the heme iron of hemoglobin and myoglobin which remains in the ferrous (Fe^{2+}) state.

Three cytochromes were initially discovered from the mammalian mitochondria. They were designated as cytochrome a, band c depending on the type of heme present and the respective absorption spectrum. Additional cytochromes such as c_1, b_l ,b_2, a_3 etc. were discovered later.

The **electrons** are **transported** from **coenzyme Q to cytochromes** (in the order) **b, c_1 c, a and a_3**. The property of reversible oxidation-reduction of heme iron Fe2+ ~ Fe3+ present in cytochromes allows them to function as effective carriers of electrons in He.

Cytochrome c (mol. wt. 13,000) is a small protein containing 104 amino acids and a heme group. It is a central member of ETC with an intermediate redox potential. It is rather loosely bound to inner mitochondrial membrane and can be easily extracted.

Cytochrome a and a_3: The term cytochrome oxidase is frequently used to collectively represent cytochrome a and a3 which is the terminal component of He. Cytochrome oxidase is the only electron carrier, the heme iron of which can directly react with molecular oxygen. Besides heme (with iron), this oxidase also contains copper that undergoes oxidation-reduction ($Cu^{2+} \rightarrow Cu^+$) during the transport of electrons. In the final stage of HC, the transported electrons, the free protons and the molecular oxygen combine to produce water.

E5. OXIDATIVE PHOSPHORYLATION

The transport of electrons through the ETC is linked with the release of free energy. The process of synthesizing ATP from ADP and Pi coupled with the electron transport chain is known as oxidative phosphorylation. The complex V (Fig. 6) of the inner mitochondrial membrane is the site of oxidative phosphorylation.

P: O Ratio: The P: O ratio refers to the number of inorganic phosphate molecules utilized for ATP generation for every atom of oxygen consumed. More appropriately, P: O ratio represents the number of molecules of ATP synthesized per pair of electrons carried through ETC. The mitochondrial oxidation of NADH with a P: O ratio of 3 can be represented by the following equation:

$$NADH + H^+ + \tfrac{1}{2}\ O_2 + 3ADP + 3\ Pi \rightarrow NAD^+ + 3ATP + 4H_20$$

P: O ratio of 2 is assigned to the oxidation of $FADH_2$.
(Note: Although yet to be proved beyond doubt, some workers suggest a P: O ratio of 2.5 for NADH + H^+, and 1.5 for $FADH_2$, based on the proton translocation).

Sites of oxidative phosphorylation in ETC: There are three reactions in the HC that are exergonic to result in the synthesis of 3 ATP molecules (Fig. 7). The three sites of ATP formation in ETC are

1. Oxidation of $FMNH_2$ by coenzyme Q.
2. Oxidation of cytochrome b by cytochrome c_1.
3. Cytochrome oxidase reaction.

Each one of the above reactions represents a coupling site for ATP production. There are only two coupling sites for the oxidation of $FADH_2$ (P: O ratio 2), since the first site is bypassed.

Energetics of oxidative phosphorylation: The transport of electrons from redox pair NAD^+/NADH (E_o =-0.32) to finally the redox pair ½ O_2/H_2O (E_o = +0.82) may be simplified and represented in the following equation

$$\tfrac{1}{2}\ O_2 + NADH + H^+ \rightarrow H_2O + NAD+$$

The redox potential difference between these two redox pairs is 1.14 V, which is equivalent to an energy 52 Cal/mol. Three ATP are synthesized in the ETC when NADH is oxidized which equals to 21.9 Cal (each ATP = 7.3 Cal). The efficiency of energy conservation is calculated as

$$\frac{21.9 \times 100}{52} = 42\%.$$

Therefore, when NADH is oxidized, about 42% of energy is trapped in the form of 3 ATP and the remaining is lost as heat.

MECHANISM OF OXIDATIVE PHOSPHORYLAT'ON

Several hypotheses have been put forth to explain the process of oxidative phosphorylation. The most important among them-namely, chemical coupling, conformational coupling and chemiosmotic-are discussed below.

Chemical coupling hypothesis: This hypothesis was put forth by Edward Slater (1953). According to chemical coupling hypothesis, during the course of electron transfer in respiratory chain, a series of phosphorylated high-energy

intermediates are first produced which are utilized for the synthesis of ATP. These reactions are believed to be analogous to the substrate level phosphorylation that occurs in glycolysis or citric acid cycle. However, this hypothesis lacks experimental evidence, since all attempts, so far, to isolate anyone of the high-energy intermediates have not been successful.

Conformational coupling hypothesis: Paul Boyer proposed this hypothesis in 1964. As a consequence of electron transport, a conformational change in the membrane protein is induced. This is believed to be responsible for the synthesis of ATP. The protein attains a high-energy conformation due to electron transfer which, when reverted back to random state (low-energy), results in the production of ATP from ADP + Pi (Fig. 8). There is some evidence demonstrating conformational changes in mitochondrial membrane proteins. However, there is no conclusive proof that such changes may be actually responsible for ATP synthesis.

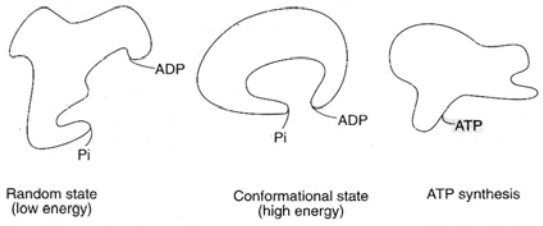


Fig. 8 : Diagrammatic representation of the conformational coupling hypothesis.

Chemiosmotic hypothesis: This mechanism, originally proposed by Peter Mitchell (1961), is now widely accepted. It explains how the transport of electrons through the respiratory chain is effectively utilized to produce ATP from ADP + Pi. The concept of chemiosmotic hypothesis is comparable with energy stored in a battery separated by positive and negative charges.

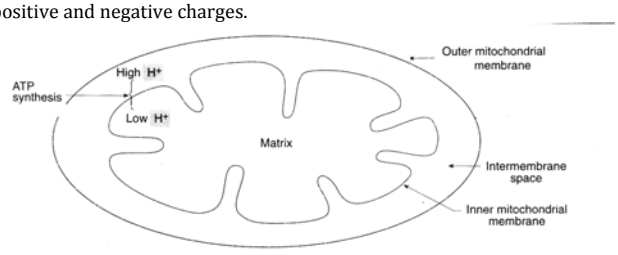


Fig. 9 : Outline of chemiosmotic hypothesis for oxidative phosphorylation.

Proton gradient: The inner mitochondrial membrane, as such, is impermeable to protons (H^+) and hydroxyl ions (OH^-). The transport of electrons through ETC is coupled with the translocation of protons (H^+) across the inner mitochondrial membrane (coupling membrane) from the matrix to the intermembrane space. The pumping of protons results in an electrochemical or proton gradient. This is due to the accumulation of more H^+ ions (low pH) on the outer side of the inner mitochondrial membrane than the inner side. The proton gradient developed due to the electron flow in the respiratory chain is sufficient to result in the synthesis of ATP from ADP and Pi.

Oxidation/reduction (o/r) loops or redox loops: It is postulated that the ETC is folded into 3 oxidation/reduction loops corresponding to complex I, III, and IV. Each loop is a coupling site and consists of a hydrogen carrier and an electron carrier. **Six protons are translocated for each electron pair transferred from NADH+H$^+$ to oxygen**. The three o/r loops, along with the carriers of protons and electrons, are depicted in Fig.11. The first pair of protons ($2H^+$) are translocated via FMN and the second ($2H^+$) via coenzyme Q.

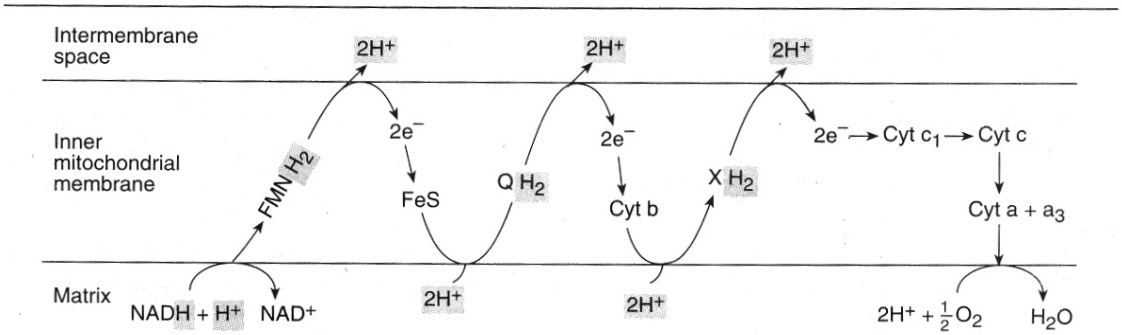


Fig. 11 : ETC with redox loops (tentative) in chemiosmotic hypothesis (X—Unidentified carrier).

There is some controversy regarding the carrier of the third proton pair (designated as X in the Fig.11). Some workers believe that coenzyme Q is involved here. Coenzyme Q is a lipid-soluble small molecule and can freely move across the inner mitochondrial membrane. A cyclic process referred to as 'coenzyme Q cycle' is said to be operative for translocation of protons at two o/r loops.

Enzyme system for ATP synthesis: ATP synthase, present in the complex V, utilizes the proton gradient for the synthesis of ATP. This enzyme is also known as ATPase since it can hydrolyse ATP to ADP and Pi. ATP synthase is a complex enzyme and consists of two functional subunits, namely F_1 and F_0 (Fig. 10). Its structure is comparable with 'lollipops'.

The protons that accumulate on the intermembrane space re-enter the mitochondrial matrix leading to the synthesis of ATP. This process, involving ATP synthase, occurs in three stages

1. Subunit F_0 of ATP synthase translocates protons (re-entry) into the matrix.
2. F_1 subunit catalyses the formation of ATP from ADP and Pi. This is believed to involve a conformational change in F_1.

3. F_o and F_1, subunits coordinate with each other and couple the disappearance of proton gradient with ATP synthesis.

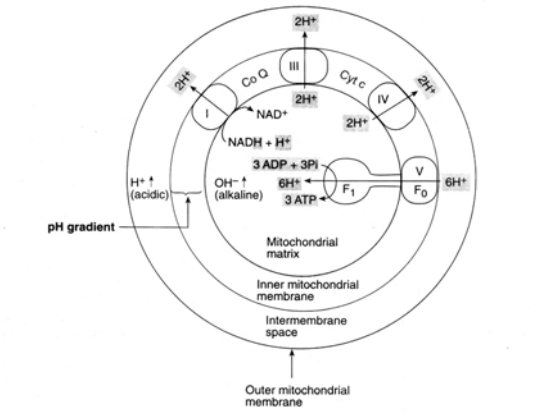


Fig. 10 : Diagrammatic representation of chemiosmotic hypothesis for oxidative phosphorylation (I, III, IV and V—Respiratory chain complexes; F_o, F_1—Protein subunits for phosphorylation).

Evidence in support of chemiosmotic hypothesis: Many experimental observations support the chemisomotic hypothesis

1. The intact inner mitochondrial membrane is responsible for oxidative phosphorylation.
2. The inner mitochondrial membrane is impermeable to protons (H^+) and other ions (OH^-, K^+, Cl^- etc.)
3. ATP can be generated by increasing the H^+ ion concentration (by adding acid) on the outside of mitochondria.
4. Certain compounds which increase the permeability of inner mitochondrial membrane to protons inhibit ATP synthesis e.g. 2,4-dinitrophenol.

Inherited disorders of oxidative phosphorylation: It is estimated that about 100 polypeptides are required for oxidative phosphorylation. Of these, 13 are coded by mitochondrial DNA (mtDNA) and synthesized in the mitochondria, while the rest are produced in the cytosol (coded by nuclear DNA) and transported. mtDNA is maternally inherited since mitochondria from the sperm do not enter the fertilized ovum.

Mitochondrial DNA is about 10 times more susceptible to mutations than nuclear DNA. mtDNA mutations are more commonly seen in tissues with high rate of oxidative phosphorylation (e.g. central nervous system, skeletal and heart muscle, liver). **Leber's hereditary optic neuropathy** is an example for mutations in mtDNA. This disorder is characterized by loss of bilateral vision due to neuroretinal degeneration.

Inhibitors of electron transport chain

Many site-specific inhibitors of ETC have contributed to the present knowledge of mitochondrial respiration. Selected examples of these inhibitors are given in Fig. 7. The inhibitors bind to one of the components of ETC and block the transport of electrons. This causes the accumulation of reduced components before the inhibitor blockade step and oxidized components after that step.The synthesis of ATP (phosphorylation) is dependent on electron transport. Hence, all the site-specific inhibitors of ETC also inhibit ATP formation. Three possible sites of action for the inhibitors of ETC are identified

1. NADH and coenzyme Q: Fish poison **rotenone, barbituate drug amytal** and **antibiotic piercidin A** inhibit at this site.

2. Between cytochrome b and c_1: **Antimycin A** -an antibiotic, **British anti lewisite** (BAL)-an antidote used against war-gas-are the two important inhibitors of the site between cytochrome b and c_1.

3. Inhibitors of cytochrome oxidase: Carbon monoxide, cyanide, hydrogen sulphide and azide effectively inhibit cytochrome oxidase. Carbon monoxide reacts with reduced form of the cytochrome while cyanide and azide react with oxidized form.

Inhibitors of oxidative phosphorylation

Uncouplers: The mitochondiral transport of electrons is tightly coupled with oxidative phosporylation (ATP synthesis). In other words, oxidation and phosphorylation proceed simultaneously. There are certain compounds that can uncouple (or delink) the electron transport from oxidative phosporylation. Such compounds, known as uncouplers, increase the permeability of inner mitochondrial membrane to protons (H^+). Thus, an uncoupler allows the electron transport but blocks the establishment of proton gradient across the inner mitochondrial membrane. The result is that ATP synthesis does not occur. The energy linked with the transport of electrons is dissipated as heat. The **uncouplers allow** (often at accelerated rate) **oxidation of substrates** (via NADH or FADH2) **without ATP formation**.

The uncoupler, **2,4-dinitrophenol (DNP)** , has been extensively studied. It is a small lipophilic molecule. DNP is a proton-carrier and can easily diffuse through the inner mitochondrial membrane. The other uncouplers include **dinitro cresol**, **penta chlorophenol**, **trifluoro carbonylcyanide phenyl hydrazone** (FCCP). The last compound (FCCP) is said to be 100 times more effective as an uncoupler than dinitrophenol. When administered in high doses, the drug aspirin acts as an uncoupler.

Physiological uncouplers: Certain physiological substances which act as uncouplers at higher concentration have been identified. These include the hormone **thyroxine** and long chain free fatty acids. The unconjugated **bilirubin** is also believed to act as an uncoupler. This is, however, yet to be proved beyond doubt.

Significance of uncoupling: Uncoupling of respiration from oxidative phosphorylation under natural conditions assumes biological significance. The maintenance of body temperature is particularly important in hairless animals, hibernating animals and the animals adapted to cold. These animals possess a specialized tissue called brown adipose tissue in the upper back and neck portions. The mitochondria of brown adipose tissue are rich in electron carriers and are specialized to carry out an oxidation uncoupled from phosphorylation. This causes liberation of heat when fat is oxidized in the brown adipose tissue. The presence of active brown adipose tissue in certain individuals is believed to protect them from becoming obese. The excess calories consumed by these people are burnt and liberated as heat, instead of being stored as fat.

Ionophores: The term 'ionophores' is used to collectively represent the lipophilic substances that promote the transport of ions across biological membranes. All the uncouplers (described above) are, in fact, proton ionophores.

The antibiotics **valinomycin** and **nigercin** act as **ionophores for K^+** ions. Both these compounds are also capable of dissipating proton gradient across the inner mitochondrial membrane and inhibit oxidative phosphorylation.

Other inhibitors of oxidative phosphorylation

Oligomycin: This antibiotic prevents the mitochondrial oxidation as well as phosphorylation. It binds with the enzyme ATP synthetase and blocks the proton (H^+) channels. It thus prevents the translocation (re-entry) of protons into the mitochondrial matrix. Due to this, protons get accumulated at higher concentration in the intermembrane space. Electron transport (respiration) ultimately stops, since protons cannot be pumped out against steep proton gradients.The action of oligomycin indicates that oxidation and phosphorylation are tightly coupled processes.

Atractyloside: This is a plant toxin and inhibits oxidative phosphorylation by an indirect mechanism. The presence of adequate concentration of ADP and Pi in the mitochondrial matrix and the transport of ATP into the extra-mitochondrial system are needed for phosphorylation to proceed continuously. Adenine nucleotide carrier system facilitates the transport of ATP and ADP. Atractyloside inhibits adenine nucleotide carrier and, thus, **blocks the adequate supply of ADP**, thereby preventing phosphorylation.

TRANSPORT OF REDUCING EQUIVALENTS-SHUTTLE PATHWAYS

The inner mitochondrial membrane is impermeable to NADH. Therefore, the NADH produced in the cytosol cannot directly enter the mitochondria. Two pathways-namely **glycerol phosphate shuttle** and **malate-aspartate shuttle** are operative to do this job. They transport the educing equivalents from cytosol to mitochondria and not vice versa. The shuttle pathways are not operative to the same extent in all the tissues. **Malate-aspartate shuttle** predominantly occurs in **liver and heart** whereas **glycerol-phosphate** shuttle operates in **muscle and brain**.

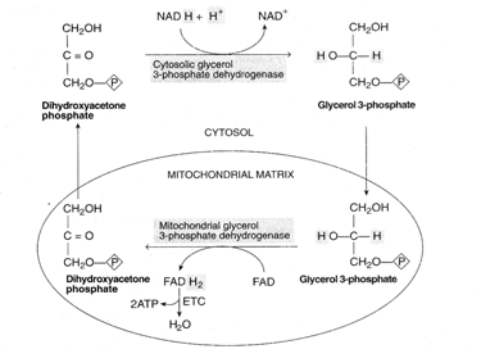


Fig. 12 : Glycerol-phosphate shuttle (reducing equivalents transported are shown in green shade).

I. Glycerol-phosphate shuttle: Cytosolic glycerol 3-phosphate dehydrogenase oxidizes NADH to NAD^+. The reducing equivalents are transported through glycerol 3-phosphate into the mitochondria. Glycerol 3-phosphate dehydrogenase-present on outer surface of inner mitochondrial membrane-reduces FAD to $FADH_2$. Dihydroxyacetone phosphate escapes into the cytosol and the shuttling continues as depicted in Fig.12. $FADH_2$ gets oxidized via ETC to generate 2 ATP.

II. Malate-aspartate shuttle: In the cytosol, oxaloacetate accepts the reducing equivalents (NADH) and becomes malate. Malate then enters mitochondria where it is oxidized by mitochondrial malate dehydrogenase. In this reaction, NADH and oxaloacetate are regenerated. NADH gets oxidized via electron transport chain and 3 ATP are produced. This is in contrast to glycerol-phosphate shuttle where only 2 ATP are produced. In the mitochondria, oxaloacetate participates in transamination reaction with glutamate to produce aspartate and α-ketoglutarate. aspartate enters the cytosol and transaminates with α-ketoglutarate to give oxaloacetate and glutamate. The malate-aspartate shuttle is shown in Fig. 13.

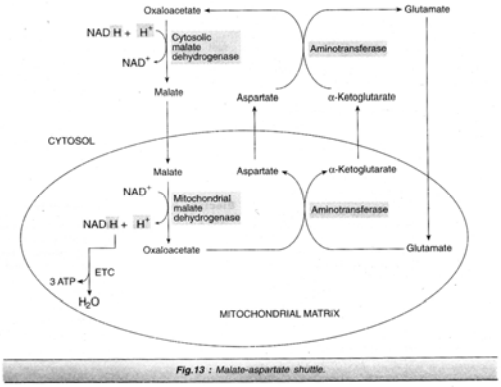


Fig.13 : Malate-aspartate shuttle.

E6. COUPLED REACTIONS AND HIGH-ENERGY COMPOUNDS (ENERGY RICH COMPOUNDS)

Metabolism is essentially a linked series of chemical reactions that begins with a particular molecule and converts it into some other molecule or molecules in a carefully defined fashion. There are many such defined pathways in the cell, and we will examine a few of them in some detail later. These pathways are interdependent, and their activity is coordinated by exquisitely sensitive means of communication in which allosteric enzymes are predominant.

We can divide metabolic pathways into two broad classes: (1) those that convert energy into biologically useful forms and (2) those that require inputs of energy to proceed. Although this division is often imprecise, it is nonetheless a useful distinction in an examination of metabolism. Those reactions that transform fuels into cellular energy are called *catabolic reactions* or, more generally, *catabolism*.

$$\text{Fuels (carbohydrates, fats)} \xrightarrow{\text{Catabolism}} CO_2 + H_2O + \text{useful energy}$$

Those reactions that require energy—such as the synthesis of glucose, fats, or DNA—are called *anabolic reactions* or *anabolism*. The useful forms of energy that are produced in catabolism are employed in anabolism to generate complex structures from simple ones, or energy-rich states from energy-poor ones.

$$\text{Useful energy} + \text{small molecules} \xrightarrow{\text{Anabolism}} \text{complex molecules}$$

Some pathways can be either anabolic or catabolic, depending on the energy conditions in the cell. They are referred to as *amphibolic pathways*.

A Thermodynamically Unfavorable Reaction Can Be Driven by a Favorable Reaction:

How are specific pathways constructed from individual reactions? A pathway must satisfy minimally two criteria: (1) the individual reactions must be *specific* and (2) the entire set of reactions that constitute the pathway must be *thermodynamically favored*. A reaction that is specific will yield only one particular product or set of products from its reactants. The function of enzymes is to provide this specificity. The thermodynamics of metabolism is most readily approached in terms of free energy. A reaction can occur spontaneously only if ΔG, the change in free energy, is negative. Recall that ΔG for the formation of products C and D from substrates A and B is given by

$$\Delta G = \Delta G^{\circ\prime} + RT \ln \frac{[C][D]}{[A][B]}$$

Thus, the ΔG of a reaction depends on the *nature* of the reactant and products (expressed by the $\Delta G^{\circ\prime}$ term, the standard free-energy change) and on their *concentrations* (expressed by the second term).

An important thermodynamic fact is that *the overall free-energy change for a chemically coupled series of reactions is equal to the sum of the free energy changes of the individual steps*. Consider the following reactions:

$$\begin{array}{ll} A \rightleftharpoons B + C & \Delta G^{\circ\prime} = +5 \text{ kcal mol}^{-1} (+21 \text{ kJ mol}^{-1}) \\ B \rightleftharpoons D & \Delta G^{\circ\prime} = -8 \text{ kcal mol}^{-1} (-34 \text{ kJ mol}^{-1}) \\ \hline A \rightleftharpoons C + D & \Delta G^{\circ\prime} = -3 \text{ kcal mol}^{-1} (-13 \text{ kJ mol}^{-1}) \end{array}$$

Under standard conditions, A cannot be spontaneously converted into B and C, because D*G* is positive. However, the conversion of B into D under standard conditions is thermodynamically feasible. Because free- energy changes are additive, the conversion of A into C and D has a $\Delta G^{\circ\prime}$ of -3 kcal mol^{-1} (-13 kJ mol^{-1}), which means that it can occur spontaneously under standard conditions. Thus, *a thermodynamically unfavorable reaction can be driven by a thermodynamically favorable reaction to which it is coupled*. In this example, the chemical intermediate B, common to both reactions, couples the reactions. Thus, metabolic pathways are formed by the coupling of enzyme-catalyzed reactions such that the overall free energy of the pathway is negative.

ATP Is the Universal Currency of Free Energy in Biological Systems: Just as commerce is facilitated by the use of a common currency, the commerce of the cell metabolism—is facilitated by the use of a common energy currency, *adenosine triphosphate* (ATP). Part of the free energy derived from the oxidation of foodstuffs and from light is transformed into this highly accessible molecule, which acts as the free-energy donor in most energy-requiring processes such as motion, active transport, or biosynthesis.

ATP is a nucleotide consisting of an adenine, a ribose, and a triphosphate unit. The active form of ATP is usually a complex of ATP with Mg^{2+} or Mn^{2+}. In considering the role of ATP as an energy carrier, we can focus on its triphosphate moiety. *ATP is an energy-rich molecule because its triphosphate unit contains two phosphoanhydride bonds*. A large amount of free energy is liberated when ATP is hydrolyzed to adenosine diphosphate (ADP) and orthophosphate (P_i) or when ATP is hydrolyzed to adenosine monophosphate (AMP) and pyrophosphate (PP_i).

$$ATP + H_2O \rightleftharpoons AMP + PP_i \qquad \Delta G^{\circ\prime} = -10.9 \text{ kcal mol}^{-1} (-45.6 \text{ kJ mol}^{-1})$$

$$ATP + H_2O \rightleftharpoons ADP + P_i \qquad \Delta G^{\circ\prime} = -7.3 \text{ kcal mol}^{-1} (-30.5 \text{ kJ mol}^{-1})$$

The precise D$G^{\circ\prime}$ for these reactions depends on the ionic strength of the medium and on the concentrations of Mg^{2+} and other metal ions. Under typical cellular concentrations, the actual D*G* for these hydrolyses is approximately -12 kcal mol^{-1} (-50 kJ mol^{-1}).

The free energy liberated in the hydrolysis of ATP is harnessed to drive reactions that require an input of free energy, such as muscle contraction. In turn, ATP is formed from ADP and P_i when fuel molecules are oxidized in chemotrophs or when light is trapped by phototrophs. *This ATP—ADP cycle is the fundamental mode of energy exchange in biological systems.*

Some biosynthetic reactions are driven by hydrolysis of nucleoside triphosphates that are analogous to ATP—namely, guanosine triphosphate (GTP), uridine triphosphate (UTP), and cytidine triphosphate (CTP). The diphosphate forms of these nucleotides are denoted by GDP, UDP, and CDP, and the monophosphate forms by GMP, UMP, and CMP. Enzymes can catalyze the transfer of the terminal phosphoryl group from one nucleotide to another. The phosphorylation of nucleoside monophosphates is catalyzed by a family of nucleoside monophosphate kinases. The phosphorylation of nucleoside diphosphates is catalyzed by nucleoside diphosphate kinase, an enzyme with broad specificity. It is intriguing to note that, although all of the nucleotide triphosphates are energetically equivalent, ATP is nonetheless the primary cellular energy carrier. In addition, two important electron carriers, NAD^+ and FAD, are derivatives of ATP. *The role of ATP in energy metabolism is paramount.*

ATP Hydrolysis Drives Metabolism by Shifting the Equilibrium of Coupled Reactions: How does coupling to ATP hydrolysis make possible an otherwise unfavorable reaction? Consider a chemical reaction that is thermodynamically unfavorable without an input of free energy, a situation common to many biosynthetic reactions. Suppose that the standard free energy of the conversion of compound A into compound B is +4.0 kcal mol^{-1} (+13 kJ mol^{-1}):

$$A \rightleftharpoons B \qquad \Delta G^{\circ\prime} = +4.0 \text{ kcal mol}^{-1} (+13 \text{ kJ mol}^{-1})$$

The equilibrium constant K'_{eq} of this reaction at 25°C is related to D$G^{\circ\prime}$ (in units of kilocalories per mole) by

$$K'_{eq} = [B]_{eq}/[A]_{eq} = 10^{-\Delta G^{\circ\prime}/1.36} = 1.15 \times 10^{-3}$$

Thus, net conversion of A into B cannot occur when the molar ratio of B to A is equal to or greater than 1.15×10^{-3}. However, A can be converted into B under these conditions if the reaction is coupled to the hydrolysis of ATP. The new overall reaction is

$$A + ATP + H_2O \rightleftharpoons B + ADP + P_i + H^+ \qquad \Delta G^{\circ\prime} = -3.3 \text{ kcal mol}^{-1} (-13.8 \text{ kJ mol}^{-1})$$

Its standard free-energy change of -3.3 kcal mol^{-1} (-13.8 kJ mol^{-1}) is the sum of the value of D$G^{\circ\prime}$ for the conversion of A into B [+4.0 kcal mol^{-1} (+12.6 kJ mol^{-1})] and the value of D$G^{\circ\prime}$ for the hydrolysis of ATP [-7.3 kcal mol^{-1} (-30.5 kJ mol^{-1})]. At pH 7, the equilibrium constant of this coupled reaction is

$$K'_{eq} = \frac{[B]_{eq}}{[A]_{eq}} \times \frac{[ADP]_{eq}[P_i]_{eq}}{[ATP]_{eq}} = 10^{3.3/1.36} = 2.67 \times 10^2$$

At equilibrium, the ratio of [B] to [A] is given by

$$\frac{[B]_{eq}}{[A]_{eq}} = K'_{eq}\frac{[ATP]_{eq}}{[ADP]_{eq}[P_i]_{eq}}$$

The ATP-generating system of cells maintains the [ATP]/[ADP][P_i] ratio at a high level, typically of the order of 500 M^{-1}. For this ratio,

$$\frac{[B]_{eq}}{[A]_{eq}} = 2.67 \times 10^2 \times 500 = 1.34 \times 10^5$$

which means that the hydrolysis of ATP enables A to be converted into B until the [B]/[A] ratio reaches a value of 1.34×10^5. This equilibrium ratio is strikingly different from the value of 1.15×10^{-3} for the reaction A → B in the absence of ATP hydrolysis. In other words, coupling the hydrolysis of ATP with the conversion of A into B has changed the equilibrium ratio of B to A by a factor of about 10^8.

We see here the thermodynamic essence of ATP's action as an *energy-coupling agent*. Cells maintain a high level of ATP by using oxidizable substrates or light as sources of free energy. The hydrolysis of an ATP molecule in a coupled reaction then changes the equilibrium ratio of products to reactants by a very large factor, of the order of 10^8. More generally, the hydrolysis of *n* ATP molecules changes the equilibrium ratio of a coupled reaction (or sequence of reactions) by a factor of 10^{8n}. For example, the hydrolysis of three ATP molecules in a coupled reaction changes the equilibrium ratio by a factor of 10^{24}. Thus, *a thermodynamically unfavorable reaction sequence can be converted into a favorable one by coupling it to the hydrolysis of a sufficient number of ATP molecules in a new reaction*. It should also be emphasized that A and B in the preceding coupled reaction may be interpreted very generally, not only as different chemical species. For example, A and B may represent activated and unactivated conformations of a protein; in this case, phosphorylation with ATP may be a means of conversion into an activated conformation. Such a conformation can store free energy, which can then be used to drive a thermodynamically unfavorable reaction. Through such changes in conformation, molecular motors such as myosin, kinesin, and dynein convert the chemical energy of ATP into mechanical energy. Indeed, this conversion is the basis of muscle contraction.

Alternatively, A and B may refer to the concentrations of an ion or molecule on the outside and inside of a cell, as in the active transport of a nutrient. The active transport of Na^+ and K^+ across membranes is driven by the phosphorylation of the sodium-potassium pump by ATP and its subsequent dephosphorylation.

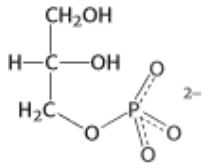
Structural Basis of the High Phosphoryl Transfer Potential of ATP

As illustrated by molecular motors and ion pumps, phosphoryl transfer is a common means of energy coupling. Furthermore, phosphoryl transfer is also widely used in the intracellular transmission of information. What makes ATP a particularly efficient phosphoryl-group donor? Let us compare the standard free energy of hydrolysis of ATP with that of a phosphate ester, such as glycerol 3-phosphate:

$$ATP + H_2O \rightleftharpoons ADP + P_i + H^+ \qquad \Delta G^{\circ\prime} = -7.3 \text{ kcal mol}^{-1} (-30.5 \text{ kJ mol}^{-1})$$

$$\text{Glycerol 3-phosphate} + H_2O \rightleftharpoons \text{glycerol} + P_i \qquad \Delta G^{\circ\prime} = -2.2 \text{ kcal mol}^{-1} (-9.2 \text{ kJ mol}^{-1})$$

The magnitude of $\Delta G^{\circ\prime}$ for the hydrolysis of glycerol 3-phosphate is much smaller than that of ATP, which means that ATP has a stronger tendency to transfer its terminal phosphoryl group to water than does glycerol 3-phosphate. In other words, ATP has a higher *phosphoryl transfer potential (phosphoryl-group transfer potential)* than does glycerol 3-phosphate.



Glycerol 3-phosphate

What is the structural basis of the high phosphoryl transfer potential of ATP? Because $\Delta G^{\circ\prime}$ depends on the *difference* in free energies of the products and reactants, the structures of both ATP and its hydrolysis products, ADP and P_i, must be examined to answer this question. Three factors are important: *resonance stabilization, electrostatic repulsion*, and *stabilization due to hydration*. ADP and, particularly, P_i, have greater resonance stabilization than does ATP. Orthophosphate has a number of resonance forms of similar energy (Figure), whereas the g-phosphoryl group of ATP has a smaller number.

Forms are unfavorable because a positively charged oxygen atom is adjacent to a positively charged phosphorus atom, an electrostatically unfavorable juxtaposition. Furthermore, at pH 7, the triphosphate unit of ATP carries about four negative charges. These charges repel one another because they are in close proximity. The repulsion between them is reduced when ATP is hydrolyzed. Finally, water can bind more effectively to ADP and P_i than it can to the phosphoanhydride part of ATP, stabilizing the ADP and P_i by hydration.

ATP is often called a high-energy phosphate compound, and its phosphoanhydride bonds are referred to as high-energy bonds. Indeed, a "squiggle" (~P) is often used to indicate such a bond. Nonetheless, there is nothing special about the bonds themselves. *They are high-energy bonds in the sense that much free energy is released when they are hydrolyzed*, for the aforegiven reasons

ATP-the most important high-energy compound: Adenosine triphosphate (ATP) is a unique and the most important high-energy molecule in the living cells. It consists of an adenine, a ribose and a triphosphate moiety. ATP is a high-energy compound due to the presence of two phosphoanhydride bonds in the triphosphate unit. ATP serves as the energy currency of the cell as is evident from the ATP-ADP cycle.

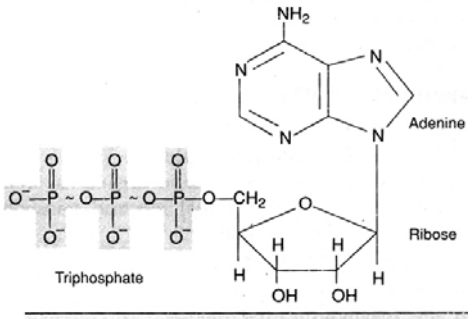


Fig. 1 : Structure of ATP.

ATP-ADP Cycle: The hydrolysis of ATP is associated with the release of large amount of energy.

$$ATP + H_2O \rightarrow \quad 7\ ADP + Pi + 7.3\ Cal.$$

The energy liberated is utilized for various processes like muscle contraction active transport etc. ATP can also act as a donor of high-energy phosphate to low-energy compounds, to make them energy rich. On the other hand, ADP can accept high-energy phosphate from the compounds possessing higher free energy content to form ATP.
ATP serves as an immediately available energy currency of the cell which is constantly being utilized and regenerated. This is represented by ATP-ADP cycle, the fundamental basis of energy exchange reactions in living system (Fig. 2). The turnover of ATP is very high. It is estimated that a resting man consumes about 40 kg ATP per day. As much as 0.5 kg ATP/minute is believed to be spent during strenous exercise. ATP acts as an **energy link between the catabolism (degradation of molecules) and anabolism (synthesis) in the biological system**.

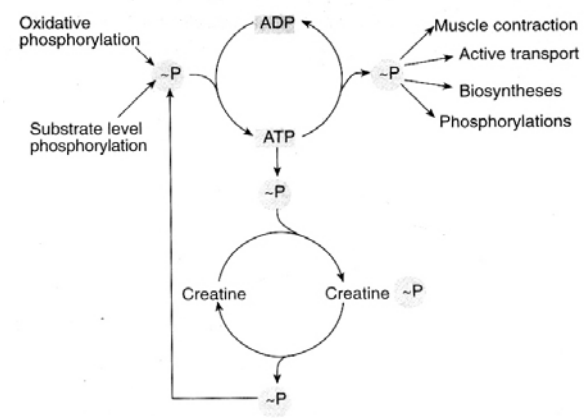


Fig. .2 : ATP-ADP cycle along with sources and utilization of ATP (Note that ~P does not exist in free form, but is only transferred).

Synthesis of ATP: ATP can be synthesized in two ways
1. Oxidative phosphorylation: This is the major source of ATP in aerobic organisms. It is linked with the mitochondrial electron transport chain (details described later).
2. Substrate level phosphorylation: ATP may be directly synthesized during substrate oxidation in the metabolism. The high-energy compounds such as phosphoenolpyruvate and 1,3-bisphosphoglycerate (intermediates of glycolysis) and succinyl CoA (of citric acid cycle) can transfer high-energy phosphate to ultimately produce ATP.

Certain compounds are encountered in the biological system which, on hydrolysis, yield energy. The term high-energy compounds or energy rich compounds are usually applied to substances which possess sufficient free energy to liberate at least 7 Cal/mol at pH 7.0. Certain other compounds which liberate less than 7.0 Cal/mol (lower than ATP hydrolysis to ADP + Pi) are referred to as low energy compounds. The list of important high energy and low-energy metabolites is given in Table 1.

All the high-energy compounds-when hydrolysed-liberate more energy than that of ATP. These include phosphoenol pyruvate, 1, 3-bisphosphoglycerate, phosphocreatine etc. Most of the high energy compounds contain phosphate group (exception acetyl CoA) hence they are called high energy phosphate compounds.

Table 1 Standard free energy of hydrolysis of some important compounds

Compounds	*ΔG° (Cal/mol)*
High-energy phosphates	
Phosphoenol pyruvate	−14.8
Carbamoyl phosphate	−12.3
Cyclic AMP	−12.0
1, 3-Bisphosphoglycerate	−11.8
Phosphocreatine	−10.3
Acetyl phosphate	−10.3
S-Adenosylmethionine*	−10.0
Pyrophosphate	−8.0
Acetyl CoA**	−7.7
ATP ⟶ ADP + Pi	−7.3
Low-energy phosphates	
ADP ⟶ AMP + Pi	−6.6
Glucose 1-phosphate	−5.0
Fructose 6-phosphate	−3.8
Glucose 6-phosphate	−3.3
Glycerol 3-phosphate	−2.2

* Sulfonium compound
** Thioester

Classification of high-energy compounds

There are at least 5 groups of high-energy compounds.
1. Pyrophosphates e.g. ATP.
2. Acyl phosphates e.g. 1, 3-bisphosphoglycerate.
3. Enol phosphates e.g. phosphoenolpyruvate.
4. Thioesters e.g. acetyl CoA.
5. Phosphoguanidines e.g. phosphocreatine.

Table 2 gives some more details on the high energy compounds, including the high-energy bonds present in each category.

Table .2 High-energy compounds

Class	*Bond*	*Example(s)*
Pyrophosphates	—C—Ⓟ—Ⓟ	ATP, pyrophosphate
Acyl phosphates	—C(=O)—O~Ⓟ	1, 3-Bisphosphoglycerate, carbamoyl phosphate acetyl phosphate
Enol phosphates	—C(=CH)—O~Ⓟ	Phosphoenol pyruvate
Thiol esters (thioesters)	—C(=O)—O~S—	Acetyl CoA, acyl CoA
Guanidio phosphates (phosphoguanidines)	—N~P	Phosphocreatine, phosphoarginine

High-energy bonds: The high-energy compounds possess acid anhydride bonds (mostly phosphoanhydride bonds) which are formed by the condensation of two acidic groups or related compounds. These bonds are referred to as high energy bonds, since the free energy is liberated when these bonds are hydrolysed. Lipmann suggested use of the symbol ~ to represent high energy bond. For instance, ATP is written as AMP~P~P.

Storage forms of high-energy phosphates: Phosphocreatine (creatine phosphate) stored in vertebrate muscle and brain is an energy-rich compound. In invertebrates, phosphoarginine (arginine phosphate) replaces phosphocreatine.

E7. Biological Energy Transducers-Photosynthesis:

1. Photosynthesis is essentially the only mechanism of energy input in the living world. Photosynthesis (photos-light, synthesis-putting together) is an anabolic process of manufacture of organic compounds inside the chlorophyll containing cells from carbon dioxide and water with the help of sunlight as a source of energy. A simple equation of photosynthesis is as follows:

$$6CO_2 + 16H_2O \xrightarrow[\text{Radiant Energy}]{\text{Chlorophyll}} C6H_{12}O_6 + 6O_2$$

However, the function of water is to provide hydrogen for the synthesis of organic compounds.All the liberated oxygen comes from it. Therefore, the equation is modified as

$$6CO_2 + 12H_2O \xrightarrow[\text{Radiant Energy}]{\text{Chlorophyll}} C6H_{12}O_6 + 6H_2O + 6O_2$$

3. Chloroplasts is the seat of photosynthesis and is best exemplified in the higher plants. The external surface of thy-lakoids contains the photosynthetic pigments and serves the ends of light reaction. The stroma, on the other hand, is concerned with the events of the dark reaction.

4. Photosynthetic Pigments: The photosynthetjc pigments present in thylakoid membranes consist largely of two kinds of green chlorophylls, Chlorophyll a ($C_{55}H_{70}O_5N_4Mg$) and Chlorophyll b ($C_{55}H_{72}O_6N_4Mg$). Also present are yellow to orange pigments classified as carotenoids. There are two kinds of carotenoids, the pure hydrocarbon carotenes and the oxygen-containing xanthophylls. Certain carotenoids, especially violaxanthin, a xanthophyll, also exist in the chloroplast envelope, giving it a yellowish colour. In most plants, including green algae, β-carotene and lutein are the most abundant carotenoids in the thylakoids.

5. Photosynthesis consists of two types of reactions: a light dependent one and a light independent one. The light-dependent reaction is a photochemical reaction or light reaction as it came to be called, culminating in the generation of NADPH2, ATP and evolution of molecular oxygen. The NADPH2 and ATP are energy-rich, having caught the electrons that became available when light impinged upon chlorophyll. They form the assimilatory power, utilized for CO_2-fixation. The events of CO_2-fixation is light independent reaction and is designated as dark reaction.

6. Light reaction consists of two phase: **Phase I-Energy absorption** (Absorption and retention of light by the photosynthetic pigments); and **Phase II-Energy transduction** (conversion of light energy absorbed in phase I into chemical energy-ATP and $NADPH_2$ by photophosphorylation).

7. Most green plants absorb light in the visible spectrum (390-700 nm), whereas purple bacteria employ wavelengths ranging from near ultraviolet to infrared (800-950 nm). This range of the spectrum through which photosynthesis can take place is called photosynthetically active radiation. But the entire range is not employable in photosynthesis. The green plants, for example, absorb light maximally in the red and blue regions of the spectrum. A study of the absorption spectra shows the quantitative relationship between the wavelength of light and its absorption by the pigment in question. Thus, we see that chlorophyll-a has its absorption peaks at 660 nm and 430 nm chlorophyll-b at 648 nm and 456 nm carotene at 478 nm and 449 nm and xanthophyll same as carotene.

8. Photosystems: Photosystem I has been located in the thylakoid membranes. It is made up of three forms of chlorophyll-a, one absorbing maximally at 683 nm, the second absorbing maximally at 695 nm and the third at 670 nm. The last of these has been called P-700. Photosystem II had. been located in the stroma thylakoids. It is made up of two forms of chlorophyll-a with maximum absorption at 670 and 690 nm. The second form is christened P-690. Each photosystem has three components : (i) a reaction centre made up of a special chlorophyll molecule-in photosystem I it is a protein-bound chlorophyll-a molecule, P-700; in photosystem II it is P-690. The reaction centres are the actual sites where light energy is converted to chemical energy. (ii) some electron carriers-in photosystem I, X, plastocyanin, cytochrome-f and ferrodoxin as the electron carrier; photosystem II has plastoquinone and cytochrome b-559. (iii) other chlorophyll and carotenoids, which merely serve to transfer the light absorbed by them to the active centres.

Photosystem I takes part in both cyclic and non-cyclic photophosphorylations. PS-I can carryon cyclicphotosphosphorylation independently. Normally it drives an electron from photosystem II to $NADP^+$.
Photosystem II picks up electron released during photolysis of water. The same is extruded on absorption of light energy. As the extruded electron passes over cytochrome complex, sufficient energy is released to take part in the synthesis of ATP from ADP and inorganic phosphate. This photophosphorylation is noncyclic. PS II can operate only in conjunction with PS I.

9. The excited molecules of P-700. and P-690 transduce their energies to generate ATP and $NADPH_2$. Molecular oxygen is also produced but it escapes out of the photosynthetic system. ATP and $NADPH_2$ together constitute the assimilatory power and are employed in the fixation of CO_2 in the dark reaction.

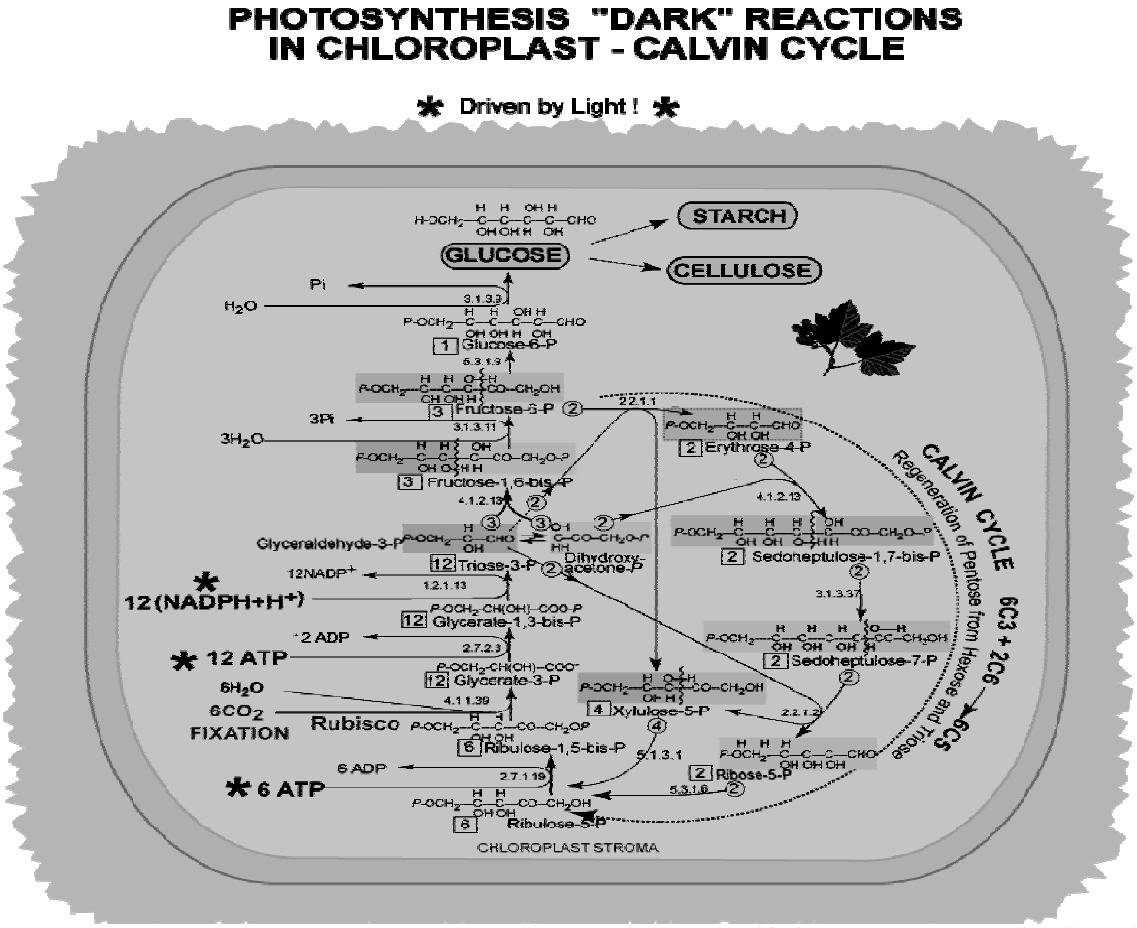
10. The oxygen that is evolved during photosynthesis comes from water and it is part of the photoact II mediated by PSII. Another aspect of oxygen evolution during photosynthesis is its relationship to the presence of certain ions in the medium such as Cl^-, Mn2+ and bicarbonate.

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^-$$

11. The production of carbon dioxide to a carbohydrate is the essence of this phase and this is accomplished through the employment of the assimilatory power (ATP and $NADPH_2$) generated in the light reaction. This energy stabilization is a dark reaction. The enzymes required for the process are present in the matrix or stroma of the chloroplast. There are two mam pathways for the biosynthetic or dark phase-Calvin cycle and C_4 dicarboxylic acid cycle. The plants exhibiting the two are respectively called C_3 and C_4 plants.

Calvin cycle

12. This cycle was discovered by Calvin, Benson and their colleagues using unicellular algae Chlorella pyrenoidosa and Scenedesmus obliques and radioactive isotope of ^{14}C with a half-life of more than 5000 years. Calvin cycle is divided into the following three phases-carboxylation, glycolytic reversal and regeneration of RuBP and are represented in diagram.



13. Photorespiration: Photorespiration is the light dependent utilization of oxygen and release of carbon dioxide by the photosynthetic organs of a plant. Normally photosynthetic organs do the reverse in the light i.e., uptake of CO_2 and release of O_2. Therefore, photorespiration is difficult to demonstrate. It is inferred from (i) Decrease in the rate of net photosynthesis when oxygen concentration is increased from 2-3% to 21% (ii) Sudden increased evolution of O_2 when an illuminated green organ is transferred to dark.

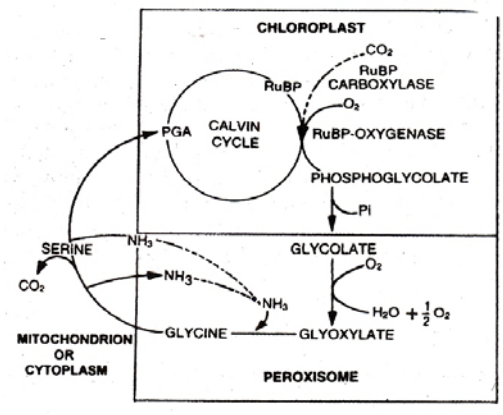


Fig. : Mechanism of photorespiration.

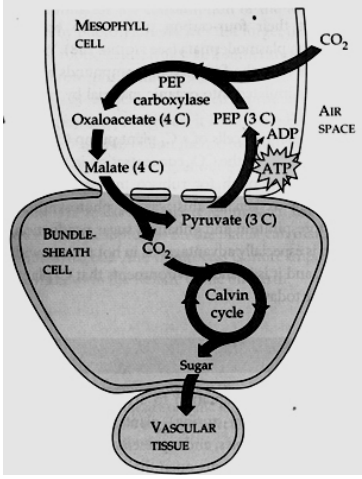
The site for photorespiration is chloroplast. Peroxisome is required for completing the process. RuBP carboxylase is changed to RuBP oxygenase. This happens at high temperature and high oxygen concentration. At high temperature and high oxygen concentration, the affinity of RuBP carboxylase for CO_2 decreases and the affinity for O_2 increases. High temperature occurs in tropical areas.

14. C4-Dicarboxylic Acid Pathway (Hatch Slack Pathway, C4 Pathway):

- Hatch and Slack found it a regular mode of CO_2-fixation in a number of tropical plants, both monocots and dicots, eg., Maize, Sugarcane, Sorghum, Panicum, Pennisetum, Atriplex, Amaranthus, Salsola etc.
- These plants are called **C4 plants** because of the first stable photosynthetic product being a 4-carbon-compound. Other plants are C_3 plants. C_4 plants often live in hot, arid and saline habitats. They have Kranz anatomy.
- In Kranz anatomy, the mesophyll is undifferentiated and its cells occur in concentric layers around vascular bundles having large bundle sheath cells. The mesophyll and bundle sheath cells are connected by plasmodesmata or cytoplasmic bridges. The chloroplasts of the mesophyll cells are smaller. They have well developed grana and a peripheral reticulum but no starch. The chloroplasts of the bundle sheath cells are larger. They have ill defined grana, a peripheral reticulum and starch grains.
- In C4 plants, initial fixation of carbon dioxide occurs in mesophyll cells. The primary acceptor of CO_2 is phosphoenol pyruvate or PEP. It combines with carbon dioxide in the presence of PEP carboxylase or pepco to form oxalo-acetic acid or oxaloacetate(4C compound).Oxalo-acetic acid is reduced to malic acid or aminated to form aspartic acid. Malic acid or aspartic acid is

translocated to bundle sheath cells through plasmodesmata. Inside the bundle sheath cells they are decarboxylated (and demainated in case of aspartic acid) to form pyruvate and CO_2.

- CO_2 is again fixed inside the bundle sheath cells through Calvin cycle. RuBP of Calvin cycle is called secondary or final acceptor of CO_2 in C4 plants. Pyruvate is sent back to mesophyll cells. Here, it is changed to phosphoenol pyruvate. Energy is required for this. The same is provided by ATP. The latter is changed into AMP (adenosine monophosphate).



Importance: The C4-plants are considered to possess greater photosynthetic efficiency, for they can utilize CO_2 until a level of 5 ppm is reached but the Calvin cycle plants cannot utilise CO_2 if the level falls below 40-50 ppm. C4-plants can utilize greater light intensities and their temperature optima for photosynthesis exceeds those of C3 plants. The chloroplasts of these plants seem to generate more ATP which of course makes for improved cellular work.The presence of extensive peripheral reticulum in the chloroplasts of these plants indirectly suggests quicker transport of products and therefore greater utilisation of light and CO_2.

15. Crassulacean Acid Metabolism (CAM):

This pathway is found in succulents, mostly members of Crassulaceae (Bryophylum and Sedum) and a few members of Bromeliaceae, such as pineapple. It is a device designed to meet the pressures of heavy transpiration, arising from their xerophytic environment. These plants obtain their CO_2 - requirements during the night time when they keep their stomata open and as CO_2 build-up occurs, the cell sap turns acidic. This process is called as dark acidification. In the following daytime, the stomata remain closed, minimizing transpirational losses, but with the advent of light, the CO_2 absorbed during the preceding night is utilized for photosynthetic purposes, the process, of light deacidification then occurs. Thus, there is a, time lag between absorption and reduction of CO_2. This arrangement helps to lessen transpirational stress but is responsible for the extremely slow growth of these plants.

16. Factors Influencing Photosynthesis

1. Carbon Dioxide: CO_2 concentration of the atmosphere is 0.03% or 300 ppm. It is a limiting factor as the available CO_2 concentration is lower than the optimum for photosynthesis. Increase in its concentration upto 0.1% (1000 ppm) increases the rate of photosynthesis in most land plants. A decline is observed beyond it.

2. Light Intensity: The light intensity at which a plant can achieve maximum amount of photosynthesis is called saturation point. Its value is 800-1000 ft candles (10% of full sunlight) in shade plants, 50-70% of full sunlight in C_3 sun plants and upto 200% of full sunlight in C4 sun plants, (e.g., Sugarcane).

3. Quality of Light: Maximum photosynthesis occurs in blue-violet and red regions of the light spectrum where most of the absorption is carried out by chlorophylls. Minimum photosynthesis occurs in the green wavelengths.

4. Temperature: It does not influence light reactions of photosynthesis but affects the enzyme controlled dark reactions. The minimum temperature at which most plants start photosynthesis is 0°-5°C but it can be as low as-20°C for lichens and 35°C for some gymnosperms. The maximum temperature at which photosynthesis can occur is 55°C in some desert plants and 75°C for hot spring algae. The optimum temperature is 10°-25°C for C3 plants and 30°-45°C for C4 plants.

5. Oxygen: Small quantity of oxygen is essential for photosynthesis except in some anaerobic bacteria. C3 plants show optimum photosynthesis at low oxygen concentration.

The possible reasons are
(i)Oxygen takes part in oxidation of photosynthetic pigments, intermediates and enzymes in the presence of strong light (photo-oxidation).
(ii) Oxygen is a strong quencher of excited state of chlorophyll.
(iii) Oxygen competes with CO_2 for reducing power. However, this effect is not known in C4 plants.
(iv) It converts RuBP-carboxylase to RuBP-oxygenase. At very high oxygen content the rate of photosynthesis begins to decline in all plants. The phenomenon is called Warburg effect.

6. Water: Water supplies H^+ and electrons for carbon dioxide fixation. However, less than 1% of the total water absorbed is utilized in photosynthesis. The rest is lost in transpiration. Even a slight increase in transpiration reduces the leaf hydration that cuts down photosynthesis by causing stomatal closure and hence decreased CO_2 absorption, loss of leaf turgidity, reduced absorption of solar radiations and decrease enzymatic activity. Thus photosynthesis is more sensitive to dehydration than any other metabolic process.

7. Air Pollutants: Dust and smoke particles present in the atmosphere reduce photosynthesis by reducing light penetration and forming a layer over the plants. Sulphur dioxide, nitrogen oxides, hydrogen fluorides and other air pollutants also decrease photosynthesis.

8. Minerals: Magnesium is a component of chlorophyll. Fe, Cu and Mn are required for synthesis of chlorophyll. Mn and Cl are linked to photolysis of water at PS II. P as phosphate is essential for ATP synthesis. Enzyme activators of photosynthesis include potassium and sulphur. Lower availability of any of these minerals reduces rate of photosynthesis.

F1. Enzymes: Basic Concepts and Principles of Catalysis

Enzymes, the catalysts of biological systems, are remarkable molecular devices that determine the patterns of chemical transformations. They also mediate the transformation of one form of energy into another. The most striking characteristics of enzymes are their *catalytic power and specificity.* Catalysis takes place at a particular site on the enzyme called the *active site. Nearly all known enzymes are proteins.* However, proteins do not have an absolute monopoly on catalysis; the discovery of catalytically active RNA molecules (Ribozymes) provides compelling evidence that RNA was an early biocatalyst.

Proteins as a class of macromolecules are highly effective catalysts for an enormous diversity of chemical reactions because of their capacity *to specifically bind a very wide range of molecules.* By utilizing the full repertoire of intermolecular forces, enzymes bring substrates together in an optimal orientation, the prelude to making and breaking chemical bonds. They catalyze reactions *by stabilizing transition states,* the highest-energy species in reaction pathways. By selectively stabilizing a transition state, an enzyme determines which one of several potential chemical reactions actually takes place.

1. Characteristics of Enzymes:

- Enzymes accelerate reactions by factors of as much as a million or more. Indeed, most reactions in biological systems do not take place at perceptible rates in the absence of enzymes. They do not change the equilibrium.
- The specificity of an enzyme is due to the precise interaction of the substrate with the enzyme. This precision is a result of the intricate three-dimensional structure of the enzyme protein.
- Like catalyst, the enzymes do not start the chemical reaction not change its equilibrium but enhance the rate of reaction. Enzyme lowers the activation energy or energy required to overcome barrier of the reaction.
- Enzymes are not imported; every cell produces its own enzymes. They may be active inside the cells (endoenzymes or intracellular enzymes, 70% in mitochondria) or outside the cell (exoenzymes eg. digestive enzymes). Amylase and protease enzymes are now added to detergent for washing and rennet tablets are used for preparation of cheese.
- Enzymes have high molecular weight-6000 kD for bacterial ferrodoxin, 250,000 for catalase, 482,000 for urease and 4600,000 for pyruvate dehdrogenase. Therefore they occur in colloidal state.
- Optimum pH for enzyme action varies for enzyme to enzyme e.g., 2 for pepsin, 6.8 for salivary amylase and 8.5 for trypsin. Most intracellular enzyme works at neutral pH.
- Enzymes are thermo labile. High temperature, denaturates enzyme. Optimum temperature for enzyme activity is 30-40^0 C for most of animals and 20-30^0 C for plants. Within limits, Q_{10}=2 to 3.
- The first ribozymes discoved by Ceck etal(1981) in Tetrahymena were self splicing group I intron. It removes introns form newly synthesized RNA. The second discovery was RNA-ase P (ribonucluease P; Altman et al (1992) which separates t-RNA from hnRNAs at their 5' end. Noller et al (1992) discovered, in prokaryotes peptidyl transferase is not a protein but a component of 23 S r-RNA.
- Lysozyme is an enzyme that hydrolyzes (breaks down) the cell walls of bacteria. Lysozyme is very common; it is found in saliva, in digestive fluids, even in body secretions. It is one of the first lines of defense against bacterial infection. Lysozyme is a small enzyme, a single polypeptide chain with both alpha helix and beta sheet regions.

F2 Six Major Classes of Enzymes:

Class	Type of reaction	Example
1. Oxidoreductases	Oxidation-reduction	Lactate dehydrogenase
2. Transferases	Group transfer	Nucleoside monophosphate kinase (NMP kinase)
3. Hydrolases	Hydrolysis reactions (transfer of functional groups to water)	Chymotrypsin
4. Lyases	Addition or removal of groups to form double bonds	Fumarase
5. Isomerases	Isomerization (intramolecular group transfer)	Triose phosphate isomerase
6. Ligases	Ligation of two substrates at the expense of ATP hydrolysis	Aminoacyl-tRNA synthetase

F3 Many Enzymes Require Cofactors for Activity

The catalytic activity of many enzymes depends on the presence of small molecules termed *cofactors*, although the precise role varies with the cofactor and the enzyme. Such an enzyme without its cofactor is referred to as an *apoenzyme*; the complete, catalytically active enzyme is called a *holoenzyme*.

$$\text{Apoenzyme} + \text{cofactor} = \text{holoenzyme}$$

Cofactors can be subdivided into two groups: metals and small organic molecules. The enzyme carbonic anhydrase, for example, requires Zn^{2+} for its activity. Glycogen phosphorylase, which mobilizes glycogen for energy, requires the small organic molecule pyridoxal phosphate (PLP).

Cofactors that are small organic molecules are called *coenzymes*. Often derived from vitamins, coenzymes can be either tightly or loosely bound to the enzyme. If tightly bound, they are called *prosthetic groups*. Loosely associated coenzymes are more like cosubstrates because they bind to and are released from the enzyme just as substrates and products are. The use of the same coenzyme by a variety of enzymes and their source in vitamins sets coenzymes apart from normal substrates, however. Enzymes that use the same coenzyme are usually mechanistically similar.

Enzyme cofactors

Cofactor	Enzyme	Cofactor	Enzyme
Coenzyme		**Metal**	
Thiamine pyrophosphate	Pyruvate dehydrogenase	Zn^{2+}	Carbonic anhydrase
Flavin adenine nucleotide	Monoamine oxidase	Zn^{2+}	Carboxypeptidase
Nicotinamide adenine dinucleotide	Lactate dehydrogenase	Mg^{2+}	*Eco*RV
Pyridoxal phosphate	Glycogen phosphorylase	Mg^{2+}	Hexokinase
Coenzyme A (CoA)	Acetyl CoA carboxylase	Ni^{2+}	Urease
Biotin	Pyruvate carboxylase	Mo	Nitrate reductase
5′-Deoxyadenosyl cobalamin	Methylmalonyl mutase	Se	Glutathione peroxidase
Tetrahydrofolate	Thymidylate synthase	Mn^{2+}	Superoxide dismutase
		K^{+}	Propionyl CoA carboxylase

F4. Water-Soluble Vitamins Function As Coenzymes

The *water-soluble vitamins include*—ascorbic acid (vitamin C) and a series known as the vitamin B complex. Ascorbate, the ionized form of ascorbic acid, serves as a reducing agent (an antioxidant). The vitamin B series comprises components of coenzymes. Note that, in all cases except vitamin C, the vitamin must be modified before it can serve its function.

Water-Soluble Vitamins

Vitamin	Coenzyme	Typical reaction type	Consequences of deficiency
Thiamine (B_1)	Thiamine pyrophosphate	Aldehyde transfer	Beriberi (weight loss, heart problems, neurological dysfunction)
Riboflavin (B_2)	Flavin adenine dinucleotide (FAD)	Oxidation-reduction	Cheliosis and angular stomatitus (lesions of the mouth), dermatitis
Pyridoxine (B_6)	Pyridoxal phosphate	Group transfer to or from amino acids	Depression, confusion, convulsions
Nicotinic acid (niacin)	Nicotinamide adenine dinucleotide (NAD^+)	Oxidation-reduction	Pellagra (dermatitis, depression, diarrhea)
Pantothenic acid	Coenzyme A	Acyl-group transfer	Hypertension
Biotin	Biotin-lysine complexes (biocytin)	ATP-dependent carboxylation and carboxyl-group transfer	Rash about the eyebrows, muscle pain, fatigue (rare)
Folic acid	Tetrahydrofolate	Transfer of one-carbon components; thymine synthesis	Anemia, neural-tube defects in development
B_{12}	5′-Deoxyadenosyl cobalamin	Transfer of methyl groups; intramolecular rearrangements	Anemia, pernicious anemia, methylmalonic acidosis
C(ascorbic acid)		Antioxidant	Scurvy (swollen and bleeding gums, subdermal hemorrhages)

F5. Enzymes May Transform Energy from One Form into Another

In many biochemical reactions, the energy of the reactants is converted with high efficiency into a different form. For example, in photosynthesis, light energy is converted into chemical-bond energy through an ion gradient. In mitochondria, the free energy contained in small molecules derived from food is converted first into the free energy of an ion gradient and then into a different currency, the free energy of adenosine triphosphate. Enzymes may then use the chemical-bond energy of ATP in many ways. The enzyme myosin converts the energy of ATP into the mechanical energy of contracting muscles. Pumps in the membranes of cells and organelles, which can be thought of as enzymes that move substrates rather than chemically altering them, create chemical and electrical gradients by using the energy of ATP to transport molecules and ions. The molecular mechanisms of these energy-transducing enzymes are being unraveled.

1. The Free-Energy Change Provides Information about the Spontaneity but Not the Rate of a Reaction

Gibbs Free Energy, or ΔG, measures amount of energy available to do work (hence "free" energy) involved in reaction; is defined in terms of both heat and entropy:

Definition: $\Delta G = \Delta H - T\Delta S$

Where

ΔG = change in free energy, a measure of useful work

ΔH = change in heat content

ΔS = change in entropy

T = absolute temperature in degrees Kelvin

Example: in burning glucose (e.g. paper), ΔH = -673 kcal, $T\Delta S$ = - 13 kcal; total ΔG = -686 kcal

The free-energy change of a reaction (ΔG) tells us if the reaction can occur spontaneously:

1. A reaction can occur spontaneously only if ΔG is negative. Such reactions are said to be exergonic.
2. A system is at equilibrium and no net change can take place if ΔG is zero.
3. A reaction cannot occur spontaneously if ΔG is positive. An input of free energy is required to drive such a reaction. These reactions are termed endergonic.

Two additional points need to be emphasized. The ΔG of a reaction depends only on the free energy of the products (the final state) minus the free energy of the reactants (the initial state). The ΔG of a reaction is independent of the path (or molecular mechanism) of the transformation. The mechanism of a reaction has no effect on ΔG. For example, the ΔG for the oxidation of glucose to CO_2 and H_2O is the same whether it occurs by combustion in vitro or by a series of enzyme-catalyzed steps in a cell. The ΔG provides no information about the rate of a reaction. A negative ΔG indicates that a reaction can occur spontaneously, but it does not signify whether it will proceed at a perceptible rate. As will be discussed shortly, the rate of a reaction depends on the free energy of activation ($\Delta G^{\ddagger}$), which is largely unrelated to the ΔG of the reaction.

2. Enzymes Alter Only the Reaction Rate and Not the Reaction Equilibrium

Because enzymes are such superb catalysts, it is tempting to ascribe to them powers that they do not have. An enzyme cannot alter the laws of thermodynamics and consequently cannot alter the equilibrium of a chemical reaction. This inability means that an enzyme accelerates the forward and reverse reactions by precisely the same factor. Consider the interconversion of A and B. Suppose that, in the absence of

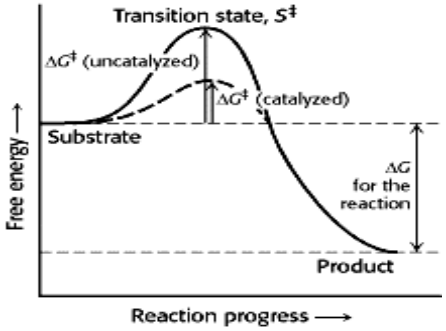
$$\mathrm{A} \underset{10^{-6}\ s^{-1}}{\overset{10^{-4}\ s^{-1}}{\rightleftharpoons}} \mathrm{B} \qquad K = \frac{[\mathrm{B}]}{[\mathrm{A}]} = \frac{k_F}{k_R} = \frac{10^{-4}}{10^{-6}} = 100$$

enzyme, the forward rate constant (kF) is 10-4 s-1 and the reverse rate constant (kR) is 10-6 s-1. The equilibrium constant K is given by the ratio of these rate constants:

The equilibrium concentration of B is 100 times that of A, whether or not enzyme is present. However, it might take considerable time to approach this equilibrium without enzyme, whereas equilibrium would be attained rapidly in the presence of a suitable enzyme. Enzymes accelerate the attainment of equilibria but do not shift their positions. The equilibrium position is a function only of the free-energy difference between reactants and products.

3 Enzymes Accelerate Reactions by Facilitating the Formation of the Transition State

The free-energy difference between reactants and products accounts for the equilibrium of the reaction, but enzymes accelerate how quickly this equilibrium is attained. A chemical reaction of substrate S to form product P goes through a transition state $S^{\ddagger}$ that has a higher free energy than does either S or P.



The double dagger denotes a thermodynamic property of the transition state. $\Delta G^{\ddagger} = G_{S^{\ddagger}} - G_S$

The transition state is the most seldom occupied species along the reaction pathway because it is the one with the highest free energy. The difference in free energy between the transition state and the substrate is called the Gibbs free energy of activation or simply the activation energy, symbolized by $\Delta G^{\ddagger}$.

Enzymes accelerate reactions by decreasing $\Delta G^{\ddagger}$, the activation energy. The combination of substrate and enzyme creates a new reaction pathway whose transition-state energy is lower than that of the reaction in the absence of enzyme. The lower activation energy means that more molecules have the required energy to reach the transition state. Decreasing the activation barrier is analogous to lowering the height of a high-jump bar; more athletes will be able to clear the bar. The essence of catalysis is specific binding of the transition state.

4. The Formation of an Enzyme-Substrate Complex Is the First Step in Enzymatic Catalysis

Much of the catalytic power of enzymes comes from their bringing substrates together in favorable orientations to promote the formation of the transition states in enzyme-substrate (ES) complexes. The substrates are bound to a specific region of the enzyme called the active site. Most enzymes are highly selective in the substrates that they bind. Indeed, the catalytic specificity of enzymes depends in part on the specificity of binding.

F6. The Active Sites of Enzymes

The active site of an enzyme is the region that binds the substrates (and the cofactor, if any). It also contains the residues that directly participate in the making and breaking of bonds. These residues are called the *catalytic groups*. In essence, *the interaction of the enzyme and substrate at the active site promotes the formation of the transition state*. The active site is the region of the enzyme that most directly lowers the $\Delta G^{=}$ of the reaction, which results in the rate enhancement characteristic of enzyme action. Although enzymes differ widely in structure, specificity, and mode of catalysis, a number of generalizations concerning their active sites can be stated:

1. *The active site is a three-dimensional cleft formed by groups that come from different parts of the amino acid sequence*—indeed, residues far apart in the sequence may interact more strongly than adjacent residues in the amino acid sequence. In lysozyme, an enzyme that degrades the cell walls of some bacteria, the important groups in the active site are contributed by residues numbered 35, 52, 62, 63, 101, and 108 in the sequence of the 129 amino acids.

2. *The active site takes up a relatively small part of the total volume of an enzyme*. Most of the amino acid residues in an enzyme are not in contact with the substrate, which raises the intriguing question of why enzymes are so big. Nearly all enzymes are made up of more than 100 amino acid residues, which gives them a mass greater than 10 kd and a diameter of more than 25 Å. The "extra" amino acids serve as a scaffold to create the three-dimensional active site from amino acids that are far apart in the primary structure. Amino acids near to one another in the primary structure are often sterically constrained from adopting the structural relations necessary to form the active site. In many proteins, the remaining amino acids also constitute regulatory sites, sites of interaction with other proteins, or channels to bring the substrates to the active sites.

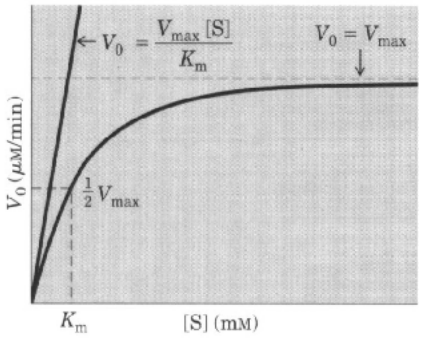
3. *Active sites are clefts or crevices*. In all enzymes of known structure, substrate molecules are bound to a cleft or crevice. Water is usually excluded unless it is a reactant. The nonpolar character of much of the cleft enhances the binding of substrate as well as catalysis. Nevertheless, the cleft may also contain polar residues. In the nonpolar microenvironment of the active site, certain of these polar residues acquire special properties essential for substrate binding or catalysis. The internal positions of these polar residues are biologically crucial exceptions to the general rule that polar residues are exposed to water.

4. *Substrates are bound to enzymes by multiple weak attractions*. ES complexes usually have equilibrium constants that range from 10^{-2} to 10^{-8} M, corresponding to free energies of interaction ranging from about -3 to -12 kcal mol^{-1} (from -13 to -50 kJ mol^{-1}). The noncovalent interactions in ES complexes are much weaker than covalent bonds, which have energies between -50 and -110 kcal mol^{-1} (between -210 and -460 kJ mol^{-1}). Electrostatic interactions, hydrogen bonds, van der Waals forces, and hydrophobic interactions mediate reversible interactions of biomolecules. Van der Waals forces become significant in binding only when numerous substrate atoms simultaneously come close to many enzyme atoms. Hence, the enzyme and substrate should have complementary shapes. The directional character of hydrogen bonds between enzyme and substrate often enforces a high degree of specificity, as seen in the RNA-degrading enzyme ribonuclease.

5. *The specificity of binding depends on the precisely defined arrangement of atoms in an active site*. Because the enzyme and the substrate interact by means of short-range forces that require close contact, a substrate must have a matching shape to fit into the site. Emil Fischer's analogy of the lock and key , expressed in 1890, has proved to be highly stimulating and fruitful. However, we now know that enzymes are flexible and that the shapes of the active sites can be markedly modified by the binding of substrate, as was postulated by Daniel E. Koshland, Jr., in 1958. The active sites of some enzymes assume a shape that is complementary to that of the transition state only *after* the substrate is bound. This process of dynamic recognition is called *induced fit*.

F7. The Michaelis-Menten Model Accounts for the Kinetic Properties of Many Enzymes

The primary function of enzymes is to enhance rates of reactions so that they are compatible with the needs of the organism. To understand how enzymes function, we need a kinetic description of their activity. For many enzymes, the rate of catalysis V_0, which is defined as the number of moles of product formed per second, varies with the substrate concentration [S]. The rate of catalysis rises linearly as substrate concentration increases and then begins to level off and approach a maximum at higher substrate concentrations.



Before we can accurately interpret this graph, we need to understand how it is generated. Consider an enzyme that catalyzes the S to P by the following pathway:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P$$

Michaelis-Menten equation:

$$V_0 = V_{max}\frac{[S]}{[S] + K_M} \qquad (23)$$

The meaning of K_m is evident from equation 23. When [S] = K_m, then $V_0 = V_{max}/2$. Thus, K_m *is equal to the substrate concentration at which the reaction rate is half its maximal value.* K_m is an important characteristic of an enzyme-catalyzed reaction and is significant for its biological function.

1. The Significance of K_M and V_{max} Values

The Michaelis constant, K_m, and the maximal rate, V_{max}, can be readily derived from rates of catalysis measured at a variety of substrate concentrations if an enzyme operates according to the simple scheme given in equation 23. The derivation of K_m and V_{max} is most commonly achieved with the use of curve-fitting programs on a computer (see the appendix to this chapter for alternative means of determining K_m and V_{max}). The K_m values of enzymes range widely. For most enzymes, K_M lies between 10^{-1} and 10^{-7} M. The K_M value for an enzyme depends on the particular substrate and on environmental conditions such as pH, temperature, and ionic strength. The Michaelis constant, K_m, has two meanings. First, K_m is the concentration of substrate at which half the active sites are filled. Thus, K_m provides a measure of the substrate concentration required for significant catalysis to occur. In fact, for many enzymes, experimental evidence suggests that K_m provides an approximation of substrate concentration in vivo. When the K_m is known, the fraction of sites filled, f_{ES}, at any substrate concentration can be calculated from

$$f_{ES} = \frac{V}{V_{max}} = \frac{[S]}{[S] + K_M} \qquad (24)$$

Second, K_m is related to the rate constants of the individual steps in the catalytic scheme. K_m is defined as $(k_{-1} + k_2)/k_1$. Consider a limiting case in which k_{-1} is much greater than k_2. Under such circumstances, the ES complex dissociates to E and S much more rapidly than product is formed. Under these conditions ($k_{-1} >> k_2$),

The dissociation constant of the ES complex is given by

$$K_M = \frac{k_{-1}}{k_1} \qquad (25)$$

$$K_{ES} = \frac{[E][S]}{[ES]} = \frac{k_{-1}}{k_1} \qquad (26)$$

In other words, K_M *is equal to the dissociation constant of the ES complex if* k_2 *is much smaller than* k_{-1}. When this condition is met, K_M is a measure of the strength of the ES complex: a high K_M indicates weak binding; a low K_M indicates strong binding. It must be stressed that K_M indicates the affinity of the ES complex only when k_{-1} is much greater than k_2.

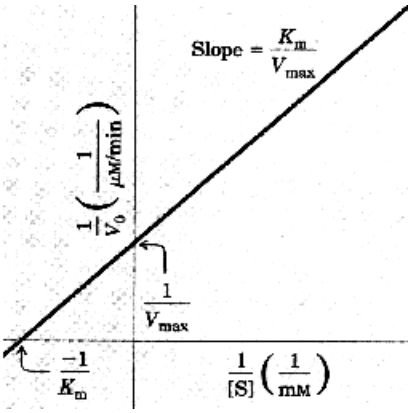
The maximal rate, V_{max}, reveals the turnover number of an enzyme, which is the number of substrate molecules converted into product by an enzyme molecule in a unit time when the enzyme is fully saturated with substrate. It is equal to the kinetic constant k_2, which is also called k_{cat}. The maximal rate, V_{max}, reveals the turnover number of an enzyme if the concentration of active sites $[E]_T$ is known, because

$$k_2 = V_{max}/[E]_T \qquad (27)$$

$$V_{max} = k_2[E]_T$$

For example, a 10^{-6} M solution of carbonic anhydrase catalyzes the formation of 0.6 M H_2CO_3 per second when it is fully saturated with substrate. Hence, k_2 is 6×10^5 s^{-1}. This turnover number is one of the largest known. Each catalyzed reaction takes place in a time equal to $1/k_2$, which is 1.7 μs for carbonic anhydrase. The turnover numbers of most enzymes with their physiological substrates fall in the range from 1 to 10^4 per second.

2. V_{max} and K_M Can Be Determined by Double-Reciprocal Plots



Before the availability of computers, the determination of K_M and V_{max} values required algebraic manipulation of the basic Michaelis-Menten equation. Because V_{max} is approached asymptotically, it is impossible to obtain a definitive value from a typical Michaelis-Menten plot. Because K_M is the concentration of substrate at $V_{max}/2$, it is likewise impossible to determine an accurate value of K_M. However, V_{max} can be accurately determined if the Michaelis-Menten equation is transformed into one that gives a straight-line plot.

Taking the reciprocal of both sides of equation 23 gives

$$\frac{1}{V_0} = \frac{1}{V_{max}} + \frac{K_M}{V_{max}} \cdot \frac{1}{[S]} \qquad (31)$$

A plot of $1/V_0$ versus 1/[S], called a *Lineweaver-Burk* or *double-reciprocal plot,* yields a straight line with an intercept of $1/V_{max}$ and a slope of K_M/V_{max} (Figure) The intercept on the *x*-axis is $-1/K_M$.

Double-reciprocal plots are especially useful for distinguishing between competitive and noncompetitive inhibitors. In competitive inhibition, the intercept on the *y*-axis of the plot of $1/V_0$ versus 1/[S] is the same in the presence and in the absence of inhibitor, although the slope is increased (Figure 1).

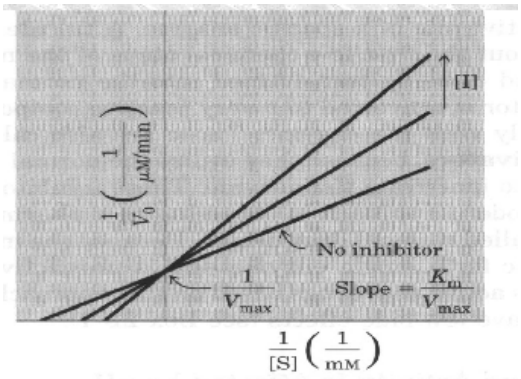


Figure 1 Competitive inhibition.

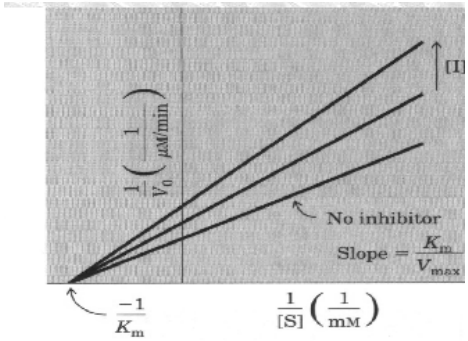


Figure 2 Noncompetitive inhibition.

That the intercept is unchanged is because a competitive inhibitor does not alter V_{max}. At a sufficiently high concentration, virtually all the active sites are filled by substrate, and the enzyme is fully operative. The increase in the slope of the $1/V_0$ versus 1/[S] plot indicates the strength of binding of competitive inhibitor. In the presence of a competitive inhibitor, equation 31 is replaced by

$$\frac{1}{V_0} = \frac{1}{V_{max}} + \frac{K_M}{V_{max}}\left(1 + \frac{[I]}{K_i}\right)\left(\frac{1}{[S]}\right)$$

in which [I] is the concentration of inhibitor and K_i is the dissociation constant of the enzyme-inhibitor complex.

E+I=EI

$K_i=[E][I]/[EI]$

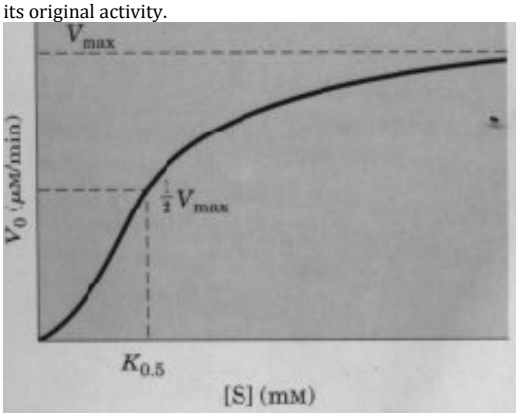
In other words, the slope of the plot is increased by the factor $(1 + [I]/K_i)$ in the presence of a competitive inhibitor. In noncompetitive inhibition, the inhibitor can combine with either the enzyme or the enzyme-substrate complex. In pure noncompetitive inhibition, the values of the dissociation constants of the inhibitor and enzyme and of the inhibitor and enzyme-substrate complex are equal. The value of V_{max} is decreased to a new value called V^{app}_{max}, and so the intercept on the vertical axis is increased. The new slope, which is equal to K_M/V^{app}_{max}, is larger by the same factor. In contrast with V_{max}, K_M is not affected by pure noncompetitive inhibition. The maximal velocity in the presence of a pure noncompetitive inhibitor, V_{max}, is given by

$$V^{app}_{max} = \frac{V_{max}}{1 + [I]/K_i}$$

F8. Allosteric Enzymes:

The Michaelis-Menten model has greatly assisted the development of enzyme chemistry. Its virtues are simplicity and broad applicability. However, the Michaelis-Menten model cannot account for the kinetic properties of many enzymes. An important group of enzymes that do not obey Michaelis-Menten kinetics comprises the allosteric enzymes. These enzymes consist of multiple subunits and multiple active sites.

The Allosteric enzymes have an Allosteric site in addition to the active site. Allosteric is derived from the greek word allo & steric (allo=other; steric=site). An effector molecule can bind with this site and can either activate or deactivate the enzyme molecule. The effector molecule is termed as modulator, regulator or ligand. The effector molecule may be either activators or inhibitors.
The activators are generally one of the substrate or cofactors. In most cases the Allosteric inhibitors are the end products of the metabolic pathways, in which particular enzyme is participating. Thus, this kind of enzyme inhibition is also called as feedback inhibition or end product inhibition or retro inhibition. The effect of Allosteric is reversible; when they are withdrawn the enzyme resumes its original activity.



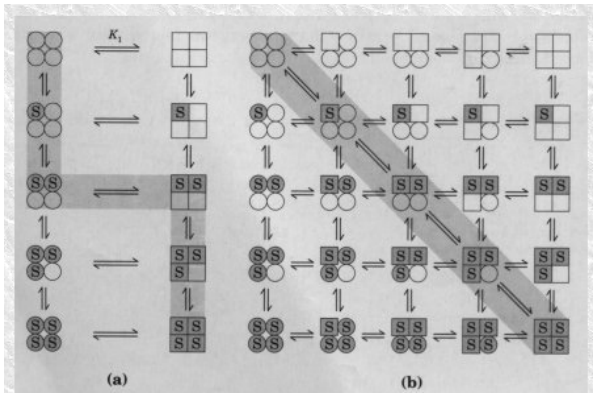
Allosteric enzymes often display sigmoidal plots of the reaction velocity V_0 versus substrate concentration [S], rather than the hyperbolic plots predicted by the Michaelis-Menten equation.

Two Models Explain the Kinetic Behavior of Allosteric Enzymes

The sigmoidal dependence of V_0 on [S] reflects subunit cooperativity, and has inspired two models to explain these cooperative interactions.

In the first model (the symmetry model), proposed by Jacques Monod and colleagues in 1965, an allosteric enzyme can exist in only two conformations, active and inactive (Fig. a). All subunits are in the active form or all are inactive. Every substrate molecule that binds increases the probability of a transition from the inactive to the active state.

In the second model (the sequential model) (Fig. b), proposed by Koshland in 1966, there are still two conformations, but subunits can undergo the conformational change individually. Binding of substrate increases the probability of the conformational change. A conformational change in one subunit makes a similar change in an adjacent subunit, as well as the binding of a second substrate molecule, more likely. There are more potential intermediate states in this model than in the symmetry model. The two models are not mutually exclusive; the symmetry model may be viewed as the "all-or-none" limiting case of the sequential model. The precise mechanism of allosteric interaction has not been established. Different allosteric enzymes may have different mechanisms for cooperative interactions



In allosteric enzymes, the binding of substrate to one active site can affect the properties of other active sites in the same enzyme molecule. A possible outcome of this interaction between subunits is that the binding of substrate becomes cooperative; that is, the binding of substrate to one active site of the enzyme facilitates substrate binding to the other active sites. Such cooperativity results in a sigmoidal plot of V_0 versus [S]. In addition, the activity of an allosteric enzyme may be altered by regulatory molecules that are reversibly bound to specific sites other than the catalytic sites. The catalytic properties of allosteric enzymes can thus be adjusted to meet the immediate needs of a cell. For this reason, allosteric enzymes are key regulators of metabolic pathways in the cell.

For eg. In threonine metabolism pathway the end product isoluecine act as inhibitor and binds to enzyme of first step, threonine deaminase.

Aspartate transcarbamoylase (ATCase), one of the best-understood allosteric enzymes, catalyzes the synthesis of N-carbamoylaspartate, the first intermediate in the synthesis of pyrimidines. ATCase is feedback inhibited by cytidine triphosphate(CTP), the final product of the pathway. ATP reverses this inhibition. ATCase consists of separable catalytic (c3) subunits (which bind the substrates) and regulatory (r2) subunits (which bind CTP and ATP). The inhibitory effect of CTP, the stimulatory action of ATP, and the cooperative binding of substrates are mediated by large

changes in quaternary structure. On binding substrates, the c3 subunits of the c6r6 enzyme move apart and reorient themselves

Positive cooperativity: Positive cooperativity means that the enzyme is much more sensitive to changes in the amount of substrate available to it. If you consider a situation such as glucose breakdown you'll realize that the amount of glucose being degraded for energy will vary greatly with the physiological situation of the cell. If it needs a great deal of energy it will be degrading a lot of glucose, if not it will be degrading very little or none at all. The enzymes in the breakdown pathway will therefore have to cope with big differences in the amount of substrate that they have to convert and the greater sensitivity to substrate concentration produced by positive substrate cooperativity will be of great assistance to this. Thus in positive cooperativity binding of substrate to any one active site increase the chance of biding substrate to other sites.

Negative cooperativity: This means that an enzyme is much less sensitive to substrate change and almost becomes independent of the concentration of substrate. This may be useful with a substrate, such as a coenzyme perhaps, whose concentration may change due to reactions in the cell which are not directly related to the pathway in which our enzyme is involved. It may well be an advantage in these circumstances if the enzyme does not react to changes in that substrate. Here binding of substrate to any one subunit decreases the affinity of enzyme for others substrates.

F9. Enzymes Can Be Inhibited by Specific Molecules (INHIBITORS)

The activity of many enzymes can be inhibited by the binding of specific small molecules and ions. This means of inhibiting enzyme activity serves as a major control mechanism in biological systems. The regulation of allosteric enzymes typifies this type of control. In addition, many drugs and toxic agents act by inhibiting enzymes. Inhibition by particular chemicals can be a source of insight into the mechanism of enzyme action: specific inhibitors can often be used to identify residues critical for catalysis. The value of transition-state analogs as potent inhibitors will be discussed shortly.

Enzyme inhibition can be either reversible or irreversible. An *irreversible inhibitor* dissociates very slowly from its target enzyme because it has become tightly bound to the enzyme, either covalently or non covalently. Some irreversible inhibitors are important drugs. Penicillin acts by covalently modifying the enzyme transpeptidase, thereby preventing the synthesis of bacterial cell walls and thus killing the bacteria. Aspirin acts by covalently modifying the enzyme cyclooxygenase, reducing the synthesis of inflammatory signals.

Reversible inhibition, in contrast with irreversible inhibition, is characterized by a rapid dissociation of the enzyme-inhibitor complex.

In *competitive inhibition*, an enzyme can bind substrate (forming an ES complex) or inhibitor (EI) but not both (ESI). The competitive inhibitor resembles the substrate and binds to the active site of the enzyme. The substrate is thereby prevented from binding to the same active site. *A competitive inhibitor diminishes the rate of catalysis by reducing the proportion of enzyme molecules bound to a substrate.* At any given inhibitor concentration, competitive inhibition can be relieved by increasing the substrate concentration. Under these conditions, the substrate "outcompetes" the inhibitor for the active site. Methotrexate is a structural analog of tetrahydrofolate, a coenzyme for the enzyme dihydrofolate reductase, which plays a role in the biosynthesis of purines and pyrimidines. It binds to dihydrofolate reductase 1000-fold more tightly than the natural substrate and inhibits nucleotide base synthesis. It is used to treat cancer.

Many antibacterial drugs work on principle of cpmpetitive inhibition for eg. sulpha drugs are believed to be competitive inhibitor of enzyme metabolizing p-amino benzoic acid(PABA). PABA is essential vitamin acting as coenzyme for many bacterial enzymatic reaction and structurally similar to sulfanilamide, a sulpha drug. Similarly malonate is structurally similar to succinate, the substrate for pyruvate dehydrogenase and thus competes for enzyme.

In *noncompetitive inhibition*, which also is reversible, the inhibitor and substrate can bind simultaneously to an enzyme molecule at different binding sites. A noncompetitive inhibitor acts by decreasing the turnover number rather than by diminishing the proportion of enzyme molecules that are bound to substrate. Noncompetitive inhibition, in contrast with competitive inhibition, cannot be overcome by increasing the substrate concentration. Such type of inhibition is shown by Allosteric enzymes which are regulated by feed back mechanism. A more complex pattern, called *mixed inhibition*, is produced when a single inhibitor both hinders the binding of substrate and decreases the turnover number of the enzyme.

1. Competitive and Noncompetitive Inhibition Are Kinetically Distinguishable

How can we determine whether a reversible inhibitor acts by competitive or noncompetitive inhibition? Let us consider only enzymes that exhibit Michaelis-Menten kinetics. Measurements of the rates of catalysis at different concentrations of substrate and inhibitor serve to distinguish the two types of inhibition. In *competitive inhibition*, the inhibitor competes with the substrate for the active site. The dissociation constant for the inhibitor is given by

$$K_i = [E][I]/[EI]$$

Because increasing the amount of substrate can overcome the inhibition, V_{max} can be attained in the presence of a competitive inhibitor. *The hallmark of competitive inhibition is that it can be overcome by a sufficiently high concentration of substrate*. However, the apparent value of K_M is altered; the effect of a competitive inhibitor is to increase the apparent value of K_M. This new value of K_M, called $K^{app}{}_M$, is numerically equal to

$$K_M^{app} = K_M(1 + [I]/K_i)$$

where [I] is the concentration of inhibitor and K_i is the dissociation constant for the enzyme-inhibitor complex. As the value of [I] increases, the value of $K^{app}{}_M$ increases. In the presence of a competitive inhibitor, an enzyme will have the same V_{max} as in the absence of an inhibitor.

In *noncompetitive inhibition*, substrate can still bind to the enzyme-inhibitor complex. However, the enzyme-inhibitor-substrate complex *does not* proceed to form product. The value of V_{max} is decreased to a new value called $V^{app}{}_{max}$ while the value of K_M is unchanged. Why is V_{max} lowered while K_M remains unchanged? In essence, the inhibitor simply lowers the concentration of functional enzyme. The remaining enzyme behaves like a more dilute solution of enzyme; V_{max} is lower, but K_M is the same. *Noncompetitive inhibition cannot be overcome by increasing the substrate concentration*.

F10. Catalytic Strategies: A Few Basic Catalytic Principles Are Used by Many Enzymes

Enzymatic catalysis begins with substrate binding. The *binding energy* is the free energy released in the formation of a large number of weak interactions between the enzyme and the substrate. We can envision this binding energy as serving two purposes: it establishes substrate specificity and increases catalytic efficiency. Only the correct substrate can participate in most or all of the interactions with the enzyme and thus maximize binding energy, accounting for the exquisite substrate specificity exhibited by many enzymes. Furthermore, the full complement of such interactions is formed only when the substrate is in the transition state. Thus, interactions between the enzyme and the substrate not only favor substrate binding but stabilize the transition state, thereby lowering the activation energy. The binding energy can also promote structural changes in both the enzyme and the substrate that facilitate catalysis, a process referred to as *induced fit.*

Enzymes commonly employ one or more of the following strategies to catalyze specific reactions:

1. *Covalent catalysis.* In covalent catalysis, the active site contains a reactive group, usually a powerful nucleophile that becomes temporarily covalently modified in the course of catalysis. The proteolytic enzyme chymotrypsin provides an excellent example of this mechanism.

2. *General acid-base catalysis.* In general acid-base catalysis, a molecule other than water plays the role of a proton donor or acceptor. Chymotrypsin uses a histidine residue as a base catalyst to enhance the nucleophilic power of serine.

3. *Metal ion catalysis.* Metal ions can function catalytically in several ways. For instance, a metal ion may serve as an electrophilic catalyst, stabilizing a negative charge on a reaction intermediate. Alternatively, the metal ion may generate a nucleophile by increasing the acidity of a nearby molecule, such as water in the hydration of CO_2 by carbonic anhydrase. Finally, the metal ion may bind to substrate, increasing the number of interactions with the enzyme and thus the binding energy. This strategy is used by NMP kinases.

4. *Catalysis by approximation.* Many reactions include two distinct substrates. In such cases, the reaction rate may be considerably enhanced by bringing the two substrates together along a single binding surface on an enzyme. NMP kinases bring two nucleotides together to facilitate the transfer of a phosphoryl group from one nucleotide to the other.

F11. Regulation of Enzymatic Activity:

The activity of proteins, including enzymes, often must be regulated so that they function at the proper time and place. The biological activity of proteins is regulated in four principal ways:

1. Allosteric control.

Allosteric proteins contain distinct regulatory sites and multiple functional sites. Regulation by small signal molecules is a significant means of controlling the activity of many proteins. The binding of these regulatory molecules at sites distinct from the active site triggers conformational changes that are transmitted to the active site.

Moreover, allosteric proteins show the property of cooperativity: activity at one functional site affects the activity at others. As a consequence, a slight change in substrate concentration can produce substantial changes in activity. Proteins displaying allosteric control are thus information transducers: their activity can be modified in response to signal molecules or to information shared among active sites.

The two best-understood allosteric proteins are: the enzyme aspartate transcarbamoylase (ATCase) and the oxygen-carrying protein hemoglobin. Catalysis by aspartate transcarbamoylase of the first step in pyrimidine biosynthesis is inhibited by cytidine triphosphate, the final product of that biosynthesis, in an example of feedback inhibition. The binding of O2 by hemoglobin is cooperative and is regulated by H+, CO2 and 2,3-bisphosphoglycerate (2,3-BPG).

2 Multiple forms of enzymes.

Isozymes, or isoenzymes, provide an avenue for varying regulation of the same reaction at distinct locations or times. Isozymes are homologous enzymes within a single organism that catalyze the same reaction but differ slightly in structure and more obviously in KM and Vmax values, as well as regulatory properties. Often, isozymes are expressed in a distinct tissue or organelle or at a distinct stage of development.

3. Reversible covalent modification.

The catalytic properties of many enzymes are markedly altered by the covalent attachment of a modifying group, most commonly a phosphoryl group. ATP serves as the phosphoryl donor in these reactions, which are catalyzed by protein kinases. The removal of phosphoryl groups by hydrolysis is catalyzed by protein phosphatases.

Cyclic AMP serves as an intracellular messenger in the transduction of many hormonal and sensory stimuli. Cyclic AMP switches on protein kinase A (PKA), a major multifunctional kinase, by binding to the regulatory subunit of the enzyme, thereby releasing the active catalytic subunits of PKA. In the absence of cAMP, the catalytic sites of PKA are occupied by pseudosubstrate sequences of the regulatory subunit

4. Proteolytic activation.

The enzymes controlled by some of these mechanisms cycle between active and inactive states. A different regulatory motif is used to irreversibly convert an inactive enzyme into an active one. Many enzymes are activated by the hydrolysis of a few or even one peptide bond in inactive precursors called zymogens or proenzymes. This regulatory mechanism generates digestive enzymes such as chymotrypsin, trypsin, and pepsin. Caspases, which are proteolytic enzymes that are the executioners in programmed cell death, or apoptosis, are proteolytically activated from the procaspase form. Blood clotting is due to a remarkable cascade of zymogen activations. Active digestive and clotting enzymes are switched off by the irreversible binding of specific inhibitory proteins that are irresistible lures to their molecular prey.

F12. Isoenzymes, Abenzymes and Riboenzymes

1. Isozymes or isoenzymes,

Enzymes that differ in amino acid sequence yet catalyze the same reaction. Usually, these enzymes display different kinetic parameters, such as KM, or different regulatory properties. They are encoded by different genetic loci, which usually arise through gene duplication and divergence. Isozymes differ from allozymes, which are enzymes that arise from allelic variation at one gene locus. Isozymes can often be distinguished from one another by biochemical properties such as electrophoretic mobility as they differ from each other in electrical charges.

The existence of isozymes permits the fine-tuning of metabolism to meet the particular needs of a given tissue or developmental stage. Consider the example of lactate dehydrogenase (LDH), an enzyme that functions in anaerobic glucose metabolism and glucose synthesis.

Human beings have two isozymic polypeptide chains for this enzyme: the H isozyme highly expressed in heart and the M isozyme found in skeletal muscle. The amino acid sequences are 75% identical. The functional enzyme is tetrameric, and many different combinations of the two subunits are possible. The H4 isozyme, found in the heart, has a higher affinity for substrates than does the M4 isozyme. The two isozymes also differ in that high levels of pyruvate, allosterically inhibit the H4 but not the M4 isozyme. The other combinations, such as H3M, have intermediate properties depending on the ratio of the two kinds of chains.

Isoenzymes are useful in separation of taxa, that is, chemotaxonomic studies and catalysis of reversible reaction, one form of enzyme can catalyze reaction in one direction, while other form of the same enzyme can catalyze the reaction in opposite direction.

2. Abzyme:

In 1969, William Jencks proposed that antibodies, specific for the transition state of a chemical reaction: should have catalytic power. This incisive prediction was realized in 1986 when the laboratories of Richard Lerner and Peter Schultz found that the catalytic antibodies could be produced by using transition state analogues as immunogens.

The preparation of an antibody that catalyze the insertion of a metal ion into a porphyrin, nicely illustrates this experimental approach. Ferrochelatase the final enzyme of bio synthetic pathway for the production of haeme, catalyze the insertion of Fe++ into porphyrin lX. The nearly planar porphyrin must be bent for iron to enter.

N-Methylprotoporphyrin is a potent inhibitor of ferrochelatase. This compound resembles the transition state because; N-alkylation forces the porphyrin to bent. Moreover, it was known that N-alkylporphyrin chelate metal ion 104 times than their unalkylated counterparts. Bending increases the exposure of lone pairs of electron of the pyrrole nitrogen to solvent, which facillates chelation.

Indeed, an effective catalyst was produced using an N-alkylporphyrin as the immunogen. The resulting antibody distorted a planar porphyrin to facilitate the entry of a metal ion. Eighty porphyrin molecules were metallated per hour antibody molecule, a rate only tenfold less than that of ferrochelatase. The uncatalysed reaction was 2500 times as slow as the antibody catalysed reaction.

Antibodies catalyzing many other kinds of chemical reactions-exemplified by ester and amide hydrolysis, amide bond formation, trans-estrification, photo-induced cleavage, photo-induced dimerization, decarboxilation and oxidization-have been produced using similar strategies. The power of transition state analogue is now evident since:

- They provide evidence insight into catalytic mechanism
- They can serve as potent and specific inhibitors of enzymes.
- They can be used as immunogens to generate a wide range of noval catalyst.

3. Ribozyme:

Thomas Cech (1983) found that the precursor of a ribosomal RNA in Tetrahymena (a ciliated protozoan) undergoes self splicing. The intron in this precursor RNA molecules is precisely removed by a catalytic activity of RNA itself. This liberated intron, then loses a short 5' terminal sequence to form a 385 nucleotide RNA molecule that catalyze the transformation of other RNA molecules. This intron-derived RNA catalyzes the cleavage and joining of RNA chains at specific site without itself being consumed. Hence, it is true enzyme- Ribozyme. Another enzyme that contains a key RNA component is ribonuclease P (RNAse P), which catalyzes the maturation of tRNA by removing nucleotides from the $5'$ end of the precursor molecule

In Tetrahymena (a ciliated protozoan), a 414-nucleotide intron is removed from a 6.4-kb precursor to yield the mature 26S rRNA molecule. In an elegant series of studies of this splicing reaction, Thomas Cech and his coworkers established that the RNA spliced itself to precisely excise the 414-nucleotide intron.

The self-splicing reaction requires an added guanosine nucleotide as an attacking group that becomes transiently incorporated into the RNA. G binds to the RNA and then attacks the $5'$ splice site to form a phosphodiester bond with the $5'$ end of the intron. This transesterification reaction generates a $3'$-OH group at the end of the upstream exon. This newly attached $3'$-OH group then attacks the $3'$ splice. This second transesterification reaction joins the two exons and leads to the release of the 414-nucleotide intron. A phosphodiester bond is formed between the two exons, and the intron is released as a linear molecule. Like catalysis by protein enzymes, self-catalysis of bond formation and breakage in this rRNA precursor is highly specific.

Messenger RNA precursors in the mitochondria of yeast and fungi also undergo self-splicing, as do some RNA precursors in the chloroplasts of unicellular organisms such as Chlamydomonas. Self-splicing reactions can be classified according to the nature of the unit that attacks the upstream splice site. Group I self-splicing is mediated by a guanosine cofactor, as in Tetrahymena. The attacking moiety in group II splicing is the $2'$-OH group of a specific adenylate of the intron.

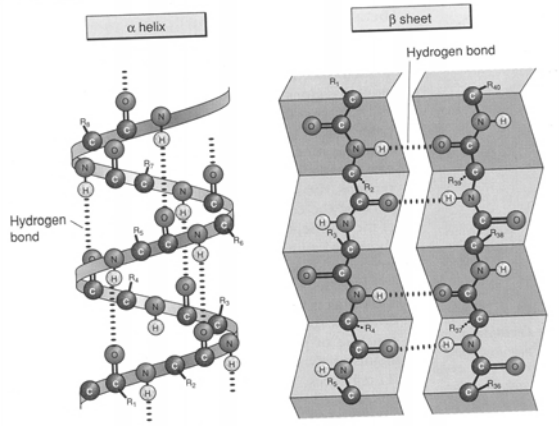
G1 CONFORMATION OF PROTEINS AND POLYPEPTIDES:

Secondary Structure:

In 1951, Linus Pauling and Robert Corey proposed two periodic structures called the α helix (alpha helix) and the β pleated sheet (beta pleated sheet). Subsequently, other structures such as the reverse turn and omega (Ω) loop were identified.

The Alpha Helix: The α helix, is a rodlike structure. A tightly coiled backbone forms the inner part of the rod and the side chains extend outward in a helical array. The α helix is stabilized by hydrogen bonds between the NH and CO groups of the main chain. In particular, the CO group of each amino acid forms a hydrogen bond with the NH group of the amino acid that is situated four residues ahead in the sequence. Thus, except for amino acids near the ends of an α helix, all the main-chain CO and NH groups are hydrogen bonded. Each residue is related to the next one by a rise of 1.5 Å along the helix axis and a rotation of 100 degrees, which gives 3.6 amino acid residues per turn of helix. Thus, amino acids spaced three and four apart in the sequence are spatially quite close to one another in an α helix. In contrast, amino acids two apart in the sequence are situated on opposite sides of the helix and so are unlikely to make contact. The pitch of the α helix, which is equal to the product of the translation (1.5 Å) and the number of residues per turn (3.6), is 5.4 Å. The screw sense of a helix can be right-handed (clockwise) or left-handed (counterclockwise). Essentially all α helices found in proteins are right-handed. In schematic diagrams of proteins, α helices are depicted as twisted ribbons or rods.

The α-helical content of proteins ranges widely, from nearly none to almost 100%. For example, about 75% of the residues in ferritin, a protein that helps store iron, are in α helices. Single α helices are usually less than 45 Å long. However, two or more α helices can entwine to form a very stable structure, which can have a length of 1000 Å (100 nm, or 0.1 μm) or more. Such α-helical coiled coils are found in myosin and tropomyosin in muscle, in fibrin in blood clots, and in keratin in hair. The helical cables in these proteins serve a mechanical role in forming stiff bundles of fibers, as in porcupine quills. The cytoskeleton (internal scaffolding) of cells is rich in so-called intermediate filaments, which also are two-stranded α-helical coiled coils. Many proteins that span biological membranes also contain α helix.

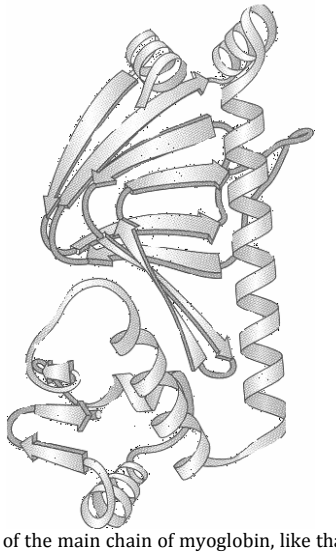


Beta Sheets: Pauling and Corey discovered another periodic structural motif, which they named the β pleated sheet. The β pleated sheet (or, more simply, the β sheet) differs markedly from the rodlike β helix. A polypeptide chain, called a β-strand, in a β sheet is almost fully extended rather than being tightly coiled as in the β helix. A range of extended structures are sterically allowed.

The distance between adjacent amino acids along a β strand is approximately 3.5 Å. The side chains of adjacent amino acids point in opposite directions. A β sheet is formed by linking two or more β strands by hydrogen bonds. Adjacent chains in a β sheet can run in opposite directions (antiparallel β sheet) or in the same direction (parallel β sheet). In the antiparallel arrangement, the NH group and the CO group of each amino acid are respectively hydrogen bonded to the CO group and the NH group of a partner on the adjacent chain. In the parallel arrangement, the hydrogen-bonding scheme is slightly more complicated. For each amino acid, the NH group is hydrogen bonded to the CO group of one amino acid on the adjacent strand, whereas the CO group is hydrogen bonded to the NH group on the amino acid two residues farther along the chain. Many strands, typically 4 or 5 but as many as 10 or more, can come together in β sheets. Such β sheets can be purely antiparallel, purely parallel, or mixed. The β sheet is an important structural element in many proteins. For example, fatty acid-binding proteins, important for lipid metabolism, are built almost entirely from β sheets.

B. Tertiary Structure:

Myoglobin, the oxygen carrier in muscle, is a single polypeptide chain of 153 amino acids. The capacity of myoglobin to bind oxygen depends on the presence of heme, a nonpolypeptide prosthetic (helper) group consisting of protoporphyrin IX and a central iron atom. Myoglobin is an extremely compact molecule. Its overall dimensions are 45 × 35 × 25 Å, an order of magnitude less than if it were fully stretched out. About 70% of the main chain is folded into eight α helices, and much of the rest of the chain forms turns and loops between helices.



The folding of the main chain of myoglobin, like that of most other proteins, is complex and devoid of symmetry. The overall course of the polypeptide chain of a protein is referred to as its tertiary structure. A unifying principle emerges from the distribution of side chains. The striking fact is that the interior consists almost entirely of nonpolar residues such as leucine, valine, methionine, and phenylalanine. Charged residues such as aspartate, glutamate, lysine, and arginine are absent from the inside of myoglobin. The only polar residues inside are two histidine residues, which play critical roles in binding iron and

oxygen. The outside of myoglobin, on the other hand, consists of both polar and nonpolar residues. The spacefilling model shows that there is very little empty space inside.

A system is more thermodynamically stable when hydrophobic groups are clustered rather than extended into the aqueous surroundings. The polypeptide chain therefore folds so that its hydrophobic side chains are buried and its polar, charged chains are on the surface. Many α helices and β strands are amphipathic; that is, the α helix or β strand has a hydrophobic face, which points into the protein interior, and a more polar face, which points into solution. The fate of the main chain accompanying the hydrophobic side chains is important, too. An unpaired peptide NH or CO group markedly prefers water to a nonpolar milieu. The secret of burying a segment of main chain in a hydrophobic environment is pairing all the NH and CO groups by hydrogen bonding. This pairing is neatly accomplished in an α helix or β sheet. Van der Waals interactions between tightly packed hydrocarbon side chains also contribute to the stability of proteins. We can now understand why the set of 20 amino acids contains several that differ subtly in size and shape. They provide a palette from which to choose to fill the interior of a protein neatly and thereby maximize van der Waals interactions, which require intimate contact.

C. Quaternary structure:

Quaternary structure refers to the spatial arrangement of subunits and the nature of their interactions. The simplest sort of quaternary structure is a dimer, consisting of two identical subunits. This organization is present in the DNA-binding protein Cro found in a bacterial virus called λ phage. More complicated quaternary structures also are common. More than one type of subunit can be present, often in variable numbers. For example, human hemoglobin, the oxygen-carrying protein in blood, consists of two subunits of one type (designated α) and two subunits of another type (designated β). Thus, the hemoglobin molecule exists as an α2β2 tetramer. Subtle changes in the arrangement of subunits within the hemoglobin molecule allow it to carry oxygen from the lungs to tissues with great efficiency.

D. Domain Structure:

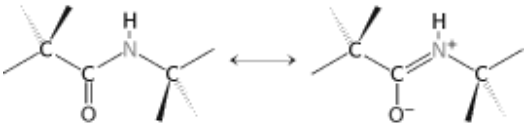
Some polypeptide chains fold into two or more compact regions that may be connected by a flexible segment of polypeptide chain, rather like pearls on a string. These compact globular units, called domains, range in size from about 30 to 400 amino acid residues. For example, the extra-cellular part of CD4, the cell-surface protein on certain cells of the immune system to which the human immunodeficiency virus (HIV) attaches itself, comprises four similar domains of approximately 100 amino acids each. Often, proteins are found to have domains in common even if their overall tertiary structures are different.

E. Reverse Turns

Most proteins have compact, globular shapes, requiring reversals in the direction of their polypeptide chains. Many of these reversals are accomplished by a common structural element called the reverse turn (also known as the β turn or hairpin bend). In many reverse turns, the CO group of residue n of a polypeptide is hydrogen bonded to the NH group of residue n + 3. This interaction stabilizes abrupt changes in direction of the polypeptide chain. In other cases, more elaborate structures are responsible for chain reversals. These structures are called loops or sometimes Ω loops (omega loops) to suggest their overall shape. Unlike α helices and β strands, loops do not have regular, periodic structures. Nonetheless, loop structures are often rigid and well defined. Turns and loops invariably lie on the surfaces of proteins and thus often participate in interactions between proteins and other molecules. The distribution of α helices, β strands, and turns along a protein chain is often referred to as its secondary structure.

G2. RAMACHANDRAN PLOT

The formation of α-helix and β- strands is at the heart of folding process. We therefore need to understand the factor that influences the secondary structure. Let us begin by considering the range of backbone conformations that are accessible to protein. Examination of the geometry of the protein backbone reveals that the peptide bond is essentially planar. Thus, for a pair of amino acids linked by a peptide bond, six atoms lie in the same plane: the α-carbon atom and CO group from the first amino acid and the NH group and α-carbon atom from the second amino acid. The nature of the chemical bonding within a peptide explains this geometric preference. The peptide bond has considerable double-bond character, which prevents rotation about this bond.



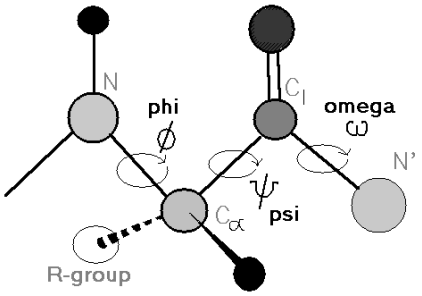
Peptide bond resonance structures

The inability of the bond to rotate constrains the conformation of the peptide backbone and accounts for the bond's planarity. This double-bond character is also expressed in the length of the bond between the CO and NH groups. The C-N distance in a peptide bond is typically 1.32 Å, which is between the values expected for a C-N single bond (1.49 Å) and a C=N double bond (1.27 Å). Finally, the peptide bond is uncharged, allowing polymers of amino acids linked by peptide bonds to form tightly packed globular structures.

Two configurations are possible for a planar peptide bond. In the trans configuration, the two α-carbon atoms are on opposite sides of the peptide bond. In the cis configuration, these groups are on the same side of the peptide bond. Almost all peptide bonds in proteins are trans. This preference for trans over cis can be explained by the fact that steric clashes between groups attached to the α-carbon atoms hinder formation of the cis form but do not occur in the trans configuration. By far the most common cis peptide bonds are X-Pro linkages. Such bonds show less preference for the trans configuration because the nitrogen of proline is bonded to two tetrahedral carbon atoms, limiting the steric differences between the trans and cis forms.

In contrast with the peptide bond, the bonds between the amino group and the α-carbon atom and between the α-carbon atom and the carbonyl group are pure single bonds. The two adjacent rigid peptide units may rotate about these bonds, taking on various orientations. This freedom of rotation about two bonds of each amino acid allows proteins to fold in many different ways. The rotations about these bonds can be specified by dihedral angles. The angle of rotation about the bond between the nitrogen and the α-carbon atoms is called phi (φ). The angle of rotation about the bond between the α-carbon and the carbonyl carbon atoms is called psi (Ψ). A clockwise rotation about either

bond as viewed from the front of the back group corresponds to a positive value. The φ and Ψ angles determine the path of the polypeptide chain.

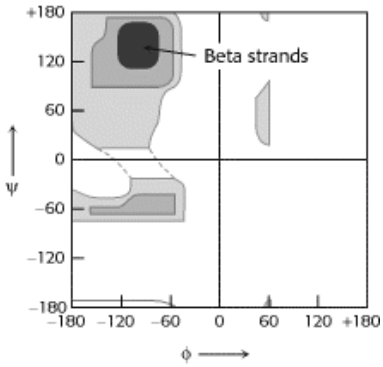


Are all combinations of φ and Ψ possible? G. N. Ramachandran recognized that many combinations are forbidden because of steric collisions between atoms. The allowed values can be visualized on a two-dimensional plot called a Ramachandran diagram. Three-quarters of the possible (φ, Ψ) combinations are excluded simply by local steric clashes. Steric exclusion, the fact that two atoms cannot be in the same place at the same time, can be a powerful organizing principle.

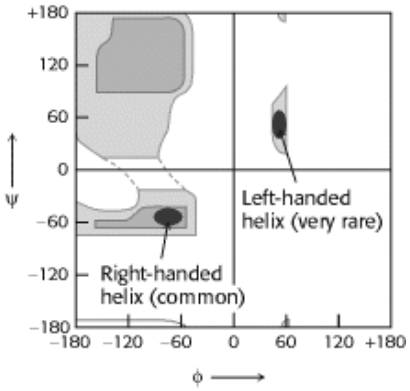
The ability of biological polymers such as proteins to fold into well defined structures is remarkable thermodynamically. Consider the equilibrium between an unfolded polymer that exists as a random coil—that is, as a mixture of many possible conformations—and the folded form that adopts a unique conformation. The favorable entropy associated with the large number of conformations in the unfolded form opposes folding and must be overcome by interactions favoring the folded form. Thus, highly flexible polymers with a large number of possible conformations do not fold into unique structures. The rigidity of the peptide unit and the restricted set of allowed φ and Ψ angles limits the number of structures accessible to the unfolded form sufficiently to allow protein folding to occur.

Repeating values of phi and psi along the chain result in regular structure. For example, repeating values of phi -57° and psi -47° give a right-handed helical fold (the alpha-helix).

The structure of cytochrome C-256 shows many segments of helix and the Ramachandran plot shows a tight grouping of phi, psi angles near to -50,-50.



Similarly, repetitive values in the region of phi = -110 to -140 and psi = +110 to +135 give extended chains with conformations that allow interactions between closely folded parallel segments (beta sheet structures). The structure of plastocyanin is composed mostly of beta sheets and the Ramachandran plot shows a broad range of values in the -110 to +130 region.



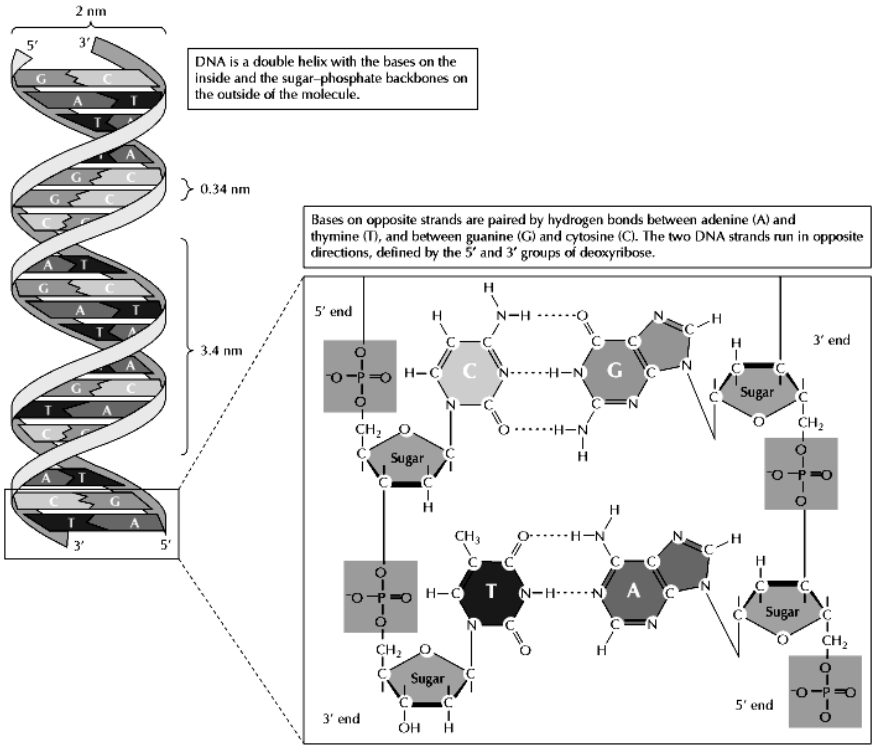
H1. STRUCTURE OF DNA AND DNA POLYMORPHISM

Utilizing X-ray diffraction data, obtained from crystals of DNA, James Watson and Francis Crick proposed a model for the structure of DNA. This model (subsequently verified by additional data) predicted that DNA would exist as a helix of two complementary antiparallel strands, wound around each other in a rightward direction and stabilized by H-bonding between bases in adjacent strands. In the Watson-Crick model, the bases are in the interior of the helix aligned at a nearly 90 degree angle relative to the axis of the helix. Purine bases form hydrogen bonds with pyrimidines, in the crucial phenomenon of base pairing.

Experimental determination has shown that, in any given molecule of DNA, the concentration of adenine (A) is equal to thymine (T) and the concentration of cytidine (C) is equal to guanine (G).

This means that A will only base-pair with T, and C with G. According to this pattern, known as Watson-Crick base-pairing, the base-pairs composed of G and C contain three H-bonds, whereas those of A and T contain two H-bonds. This makes G-C base-pairs more stable than A-T base-pairs.

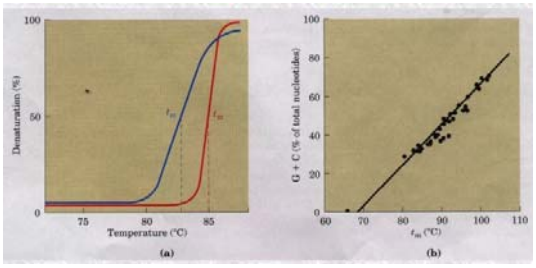
The antiparallel nature of the helix stems from the orientation of the individual strands. From any fixed position in the helix, one strand is oriented in the 5' ---> 3' direction and the other in the 3' ---> 5' direction. On its exterior surface, the double helix of DNA contains two deep grooves between the ribose-phosphate chains. These two grooves are of unequal size and termed the major and minor grooves. The difference in their size is due to the asymmetry of the deoxyribose rings and the structurally distinct nature of the upper surface of a base-pair relative to the bottom surface.



Thermal Properties of DNA:

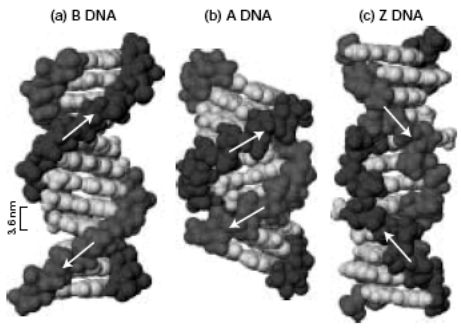
As cells divide it is a necessity that the DNA be copied (replicated), in such a way that each daughter cell acquires the same amount of genetic material. In order for this process to proceed the two strands of the helix must first be separated, in a process termed denaturation. This process can also be carried out in vitro. If a solution of DNA is subjected to high temperature, the H-bonds between bases become unstable and the strands of the helix separate in a process of thermal denaturation.

The base composition of DNA varies widely from molecule to molecule and even within different regions of the same molecule. Regions of the duplex that have predominantly A-T base-pairs will be less thermally stable than those rich in G-C base-pairs. In the process of thermal denaturation, a point is reached at which 50% of the DNA molecule exists as single strands. This point is the melting temperature (Tm), and is characteristic of the base composition of that DNA molecule. The TM depends upon several factors in addition to the base composition. These include the chemical nature of the solvent and the identities and concentrations of ions in the solution. When thermally melted DNA is cooled, the complementary strands will again re-form the correct base pairs, in a process is termed annealing or hybridization. The rate of annealing is dependent upon the nucleotide sequence of the two strands of DNA.



STRUCTURAL POLYMORPHISM OF DNA:

In addition to the major B form, three additional DNA structures have been described. Two of these are compared to B DNA in Figure. In very low humidity, the crystallographic structure of B DNA changes to the *A form;* RNA-DNA and RNA-RNA helices exist in this form in cells and in vitro. Short DNA molecules composed of alternating purinepyrimidine nucleotides (especially Gs and Cs) adopt an alternative left-handed configuration instead of the normal right-handed helix. This structure is called *Z DNA* because the bases seem to zigzag when viewed from the side. Some evidence suggests that Z DNA may occur in cells, although its function is unknown. Finally, a triple-stranded DNA structure is formed when synthetic polymers of poly(A) and polydeoxy(U) are mixed in the test tube. In addition, homopolymeric stretches of DNA composed of C and T residues in one strand and A and G residues in the other can form a triple-stranded structure by binding matching lengths of synthetic poly(C-T). Such structures probably do not occur naturally in cells but may prove useful as therapeutic agents.



By far the most important modifications in the structure of standard B-form DNA come about as a result of protein binding to specific DNA sequences. Although the multitude of hydrogen and hydrophobic bonds between the bases provide stability to DNA, the double helix is flexible about its long axis. Unlike the α helix in proteins, there are no hydrogen bonds parallel to the axis of the DNA helix. This property allows DNA to bend when complexed with a DNA-binding protein. Bending of DNA is critical to the dense packing of DNA in chromatin, the protein-DNA complex in which nuclear DNA occurs in eukaryotic cells.

Parameters of Major DNA Helices

Parameters	A Form	B Form	Z-Form
Direction of helical rotation	Right	Right	Left
Residues per turn of helix	11	10	12 base pairs
Rotation of helix per residue (in degrees)	33	36	-30
Base tilt relative to helix axis (in degrees)	20	6	7
Major groove	narrow and deep	wide and deep	Flat
Minor groove	wide and shallow	narrow and deep	narrow and deep
Orientation of N-glycosidic Bond	Anti	Anti	Anti for Py, Syn for Pu
Comments		most prevalent within cells	occurs in stretches of alternating purine-pyrimidine base pairs

H3. STRUCTURAL POLYMORPHISM OF RNA & 3D STRUCTURE OF t-RNA:

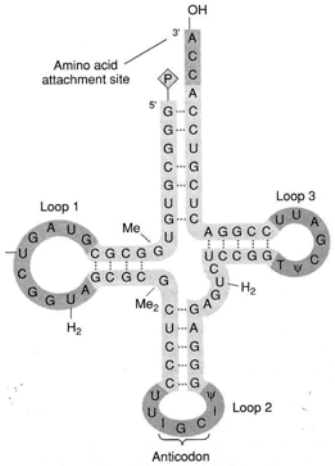
Three kinds of RNA molecules perform different but cooperative functions in protein synthesis, they are m-RNA, r-RNA and t-RNA.

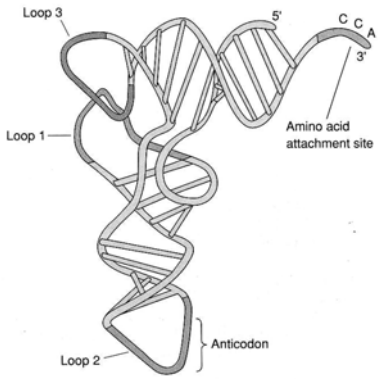
1. Messenger RNA (mRNA): It carries the genetic information copied from DNA in the form of a series of three-base code "words," each of which specifies a particular amino acid. RNA contains ribonucleotides of adenine, cytidine, guanine, and uracil; DNA contains deoxyribonucleotides of adenine, cytidine, guanine, and thymine. Because 4 nucleotides, taken individually, could represent only 4 of the 20 possible amino acids in coding the linear arrangement in proteins, a group of nucleotides is required to represent each amino acid. The code employed must be capable of specifying at least 20 words (i.e., amino acids).

Synthesis of all protein chains in prokaryotic and eukaryotic cells begins with the amino acid methionine. In most mRNAs, the start (initiator) codon specifying this aminoterminal methionine is AUG. In a few bacterial mRNAs, GUG is used as the initiator codon, and CUG occasionally is used as an initiator codon for methionine in eukaryotes. The three codons UAA, UGA, and UAG do not specify amino acids but constitute stop (terminator) signals that mark the carboxyl terminus of protein chains in almost all cells. The sequence of codons that runs from a specific start site to a terminating codon is called a reading frame. This precise linear array of ribonucleotides in groups of three in mRNA specifies the precise linear sequence of amino acids in a protein and also signals where synthesis of the protein chain starts and stops.

2. Transfer RNA (t-RNA): It is the key to deciphering the code words in mRNA. Each type of amino acid has its own type of tRNA, which binds it and carries it to the growing end of a polypeptide chain if the next code word on mRNA calls for it. The correct tRNA with its attached amino acid is selected at each step because each specific tRNA molecule contains a three-base sequence that can base-pair with its complementary code word in the mRNA. All tRNAs have two functions: to be chemically linked to a particular amino acid and to base-pair with a codon in mRNA so that the amino acid can be added to a growing peptide chain. Each tRNA molecule is recognized by one and only one of the 20 aminoacyl-tRNA synthetases. Likewise, each of these enzymes links one and only one of the 20 amino acids to a particular tRNA, forming an aminoacyl-tRNA. Once its correct amino acid is attached, a tRNA then recognizes a codon in mRNA, thereby delivering its amino acid to the growing polypeptide.As studies on tRNA proceeded, 30 ─40 different tRNAs were identified in bacterial cells and as many as 50 ─100 in animal and plant cells.

The function of tRNA molecules, which are 70 ─80 nucleotides long, depends on their precise three-dimensional structures. In solution, all tRNA molecules fold into a similar stem-loop arrangement that resembles a cloverleaf when drawn in two dimensions.





The four stems are short double helices stabilized by Watson-Crick base pairing; three of the four stems have loops containing seven or eight bases at their ends, while the remaining, unlooped stem contains the free 3′ and 5′ ends of the chain. Three nucleotides termed the anticodon, located at the center of one loop, can form base pairs with the three complementary nucleotides forming a codon in mRNA. As discussed later, specific aminoacyl-tRNA synthetases recognize the surface structure of each tRNA for a specific amino acid and covalently attach the proper amino acid to the unlooped amino acid acceptor stem. The 3′ end of all tRNAs has the sequence CCA, which in most cases is added after synthesis and processing of the tRNA are complete.

Viewed in three dimensions, the folded tRNA molecule has an L shape with the anticodon loop and acceptor stem forming the ends of the two arms.

3. Ribosomal RNA (rRNA): r-RNA associates with a set of proteins to form ribosomes. These complex structures, which physically move along an mRNA molecule, catalyze the assembly of amino acids into protein chains. They also bind tRNAs and various accessory molecules necessary for protein synthesis. Ribosomes are composed of a large and small subunit, each of which contains its own rRNA molecule or molecules.

H4. Micro RNA

In genetics, microRNAs (miRNA) are single-stranded RNA molecules of about 21–23 nucleotides in length, which regulate gene expression. miRNAs are encoded by genes that are transcribed from DNA but not translated into protein (non-coding RNA); instead they are processed from primary transcripts known as pri-miRNA to short stem-loop structures called pre-miRNA and finally to functional miRNA. Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules, and their main function is to downregulate gene expression. They were first described in 1993 by Lee and colleagues in the Victor Ambros lab, yet the term microRNA was only introduced in 2001 in a set of three articles in Science (26 October 2001). As of early 2008, computational analysis by IBM suggested the existence of as many as 50,000 different miRNAs in the typical mammalian cell, each with perhaps a thousand or more potential target.

Fig 1: *The stem-loop secondary structure of a pre-microRNA from Brassica oleracea.*

Formation and processing

The genes encoding miRNAs are much longer than the processed mature miRNA molecule; miRNAs are first transcribed as primary transcripts or pri-miRNA with a cap and poly-A tail and processed to short, 70-nucleotide stem-loop structures known as pre-miRNA in the cell nucleus. This processing is performed in animals by a protein complex known as the Microprocessor complex, consisting of the nuclease Drosha and the double-stranded RNA binding protein Pasha. These pre-miRNAs are then processed to mature miRNAs in the cytoplasm by interaction with the endonuclease Dicer, which also initiates the formation of the RNA-induced silencing complex (RISC). This complex is responsible for the gene silencing observed due to miRNA expression and RNA interference. The pathway in plants varies slightly due to their lack of Drosha homologs; instead, Dicer homologs alone effect several processing steps. The pathway is also different for miRNAs derived from intronic stem-loops; these are processed by Dicer but not by Drosha. Either the sense strand or antisense strand of DNA can function as templates to give rise to miRNA.

Efficient processing of pre-miRNA by Drosha requires the presence of extended single-stranded RNA on both 3'- and 5'-ends of hairpin molecule. These ssRNA motifs could be of different composition while their length is of high importance if processing is to take place at all. A bioinformatics analysis of human and fly pri-miRNAs revealed very similar structural regions, called 'basal segments', 'lower stems', 'upper stems' and 'terminal loops'; based on these conserved structures, thermodynamic profiles of pri-miRNA have been determined. The Drosha complex cleaves RNA molecule ~2 helical turns away from the terminal loop and ~1 turn away from basal segments. In most analysed molecules this region contains unpaired nucleotides and the free energy of the duplex is relatively high compared to lower and upper stem regions. Most pre-miRNAs don't have a perfect double-stranded RNA (dsRNA) structure topped by a terminal loop. There are few possible explanations for such selectivity. One could be that dsRNAs longer than 21 base pairs activate interferon response and anti-viral machinery in the cell. Another plausible explanation could be that the thermodynamic profile of pre-miRNA determines which strand will be incorporated into Dicer complex. Indeed, clear similarities between pri-miRNAs encoded in respective (5'- or 3'-) strands have been demonstrated.

When Dicer cleaves the pre-miRNA stem-loop, two complementary short RNA molecules are formed, but only one is integrated into the RISC complex. This strand is known as the guide strand and is selected by the argonaute protein, the catalytically active RNase in the RISC complex,

on the basis of the stability of the 5' end. The remaining strand, known as the anti-guide or passenger strand, is degraded as a RISC complex substrate. After integration into the active RISC complex, miRNAs base pair with their complementary mRNA molecules and induce mRNA degradation by argonaute proteins, the catalytically active members of the RISC complex. It is as yet unclear how the activated RISC complex locates the mRNA targets in the cell, though it has been shown that the process is not coupled to ongoing protein translation from the mRNA.

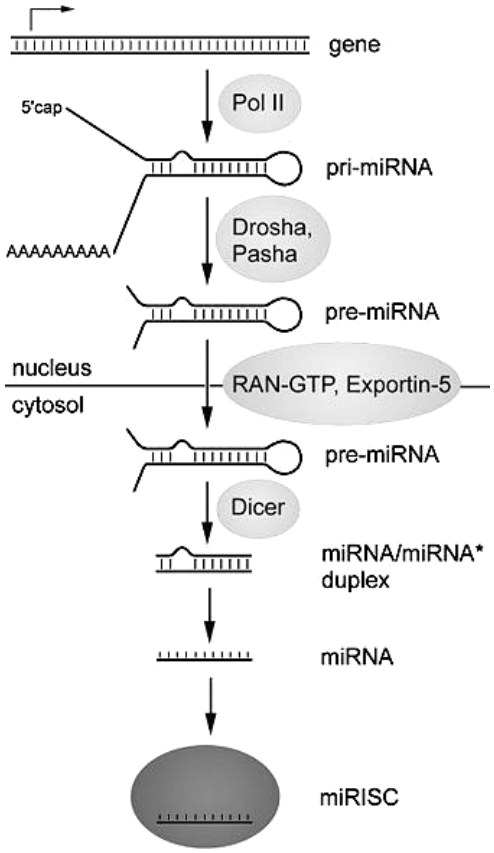


Fig 2: *MicroRNA (miRNA) is produced from precursor microRNA (pre-miRNA), which in turn is formed from a microRNA primary transcript (pri-miRNA).*

Cellular functions

The function of miRNAs appears to be in gene regulation. For that purpose, a miRNA is complementary to a part of one or more messenger RNAs (mRNAs). Animal miRNAs are usually complementary to a site in the 3' UTR whereas plant miRNAs are usually complementary to coding regions of mRNAs. The annealing of the miRNA to the mRNA then inhibits protein translation, but sometimes facilitates cleavage of the mRNA. This is thought to be the primary mode of action of plant miRNAs. In such cases, the formation of the double-stranded RNA through the binding of the miRNA triggers the degradation of the mRNA transcript through a process similar to RNA interference (RNAi), though in other cases it is believed that the miRNA complex blocks the protein translation machinery or otherwise prevents protein translation without causing the mRNA to be degraded. miRNAs may also target methylation of genomic sites which correspond to targeted mRNAs. miRNAs function in association with a complement of proteins collectively termed the miRNP.

This effect was first described for the worm C. elegans in 1993 by Victor Ambros and coworkers. As of 2002, miRNAs have been confirmed in various plants and animals, including C. elegans, human and the plant Arabidopsis thaliana. Work at the University of Louisville has resulted in the production of microarrays containing all known (at the time of production) miRNAs for human, mouse, rat, dog, C. elegans and Drosophila species, tools referred to as MMChips. Agilent has subsequently commercialized a human miRNA microarray.

Genes have been found in bacteria that are similar to eukaryotic miRNA genes in the sense that they control mRNA abundance or translation by binding an mRNA by base pairing, however they are not generally considered to be miRNAs because the Dicer enzyme is not involved.

In plants, similar RNA species termed short-interfering RNAs siRNAs are used to prevent the transcription of viral RNA. While this siRNA is double-stranded, the mechanism seems to be closely related to that of miRNA, especially taking the hairpin structures into account. siRNAs are also used to regulate cellular genes, as miRNAs do.

Gene activation

dsRNA can also activate gene expression, a mechanism that has been termed "small RNA-induced gene activation" or RNAa. dsRNAs targeting gene promoters can induce potent transcriptional activation of associated genes. This was demonstrated in human cells using synthetic dsRNAs termed small activating RNAs (saRNAs), but has also been demonstrated for endogenous microRNA.

Detecting and manipulating miRNA signaling

The activity of an miRNA can be experimentally blocked using a locked nucleic acid oligo, a Morpholino oligo or a 2'-O-methyl RNA oligo. Steps in the maturation of miRNAs can be blocked by steric-blocking oligos. The target site of an miRNA on an mRNA can be blocked by a steric blocking oligo.

miRNA and disease

Just as miRNA is involved in the normal functioning of eukaryotic cells, so has dysregulation of miRNA been associated with disease. Disease association in turn has led to increased funding opportunities for academic research and financial incentives for development and commercialization of miRNA-based diagnostics and therapeutics. After early commercialization aimed at academic research support was established, the initial research focus based on products and services requested was on cancer and neuroscience research. During 2007, interests indicated by product and services requested broadened to include cardiac research, virology, cell biology in general and plant biology.

miRNA and cancer

Several miRNAs has been found to have links with some types of cancer.

A study of mice altered to produce excess c-myc — a protein implicated in several cancers — shows that miRNA has an effect on the development of cancer. Mice that were engineered to produce a surplus of types of miRNA found in lymphoma cells developed the disease within 50 days and died two weeks later. In contrast, mice without the surplus miRNA lived over 100 days.

Another study found that two types of miRNA inhibit the E2F1 protein, which regulates cell proliferation. miRNA appears to bind to messenger RNA before it can be translated to proteins that switch genes on and off.

By measuring activity among 217 genes encoding miRNA, patterns of gene activity that can distinguish types of cancers can be discerned. miRNA signatures may enable classification of cancer. This will allow doctors to determine the original tissue type which spawned a cancer and to be able to target a treatment course based on the original tissue type. miRNA profiling has already been able to determine whether patients with chronic lymphocytic leukemia had slow growing or aggressive forms of the cancer. In 2008, the companies Asuragen and Exiqon were working to commercialize this potential for miRNAs to act as cancer biomarkers.

miRNA and heart disease

The global role of miRNA function in the heart has been addressed by conditionally inhibiting miRNA maturation in the murine heart, and has revealed that miRNAs play an essential role during its development. miRNA expression profiling studies demonstrate that expression levels of specific miRNAs change in diseased human hearts, pointing to their involvement in cardiomyopathies. Furthermore, studies on specific miRNAs in animal models have identified distinct roles for miRNAs both during heart development and under pathological conditions, including the regulation of key factors important for cardiogenesis, the hypertrophic growth response, and cardiac conductance. In 2008, academic work on the relationship between miRNA and heart disease had advanced sufficiently to lead to the establishment of a company, miRagen, with a primary focus on "cardiovascular health and disease".

I1. STABILITY OF PROTEINS

The unique three-dimensional structure of each polypeptide is determined by its amino acid sequence. Interactions between the amino acid side chains guide the folding of the polypeptide to form a compact structure. Four types of interactions cooperate in stabilizing the tertiary structures of globular proteins.

1. Disulfide bonds: A disulfide bond is a covalent linkage formed from the sulfhydryl group (-SH) of each of two cysteine residues, to produce a cystine residue (Figure 1). The two cysteines may be separated from each other by many amino acids in the primary sequence of a polypeptide, or may even be located on two different polypeptide chains; the folding of the polypeptide chain(s) brings the cysteine residues into proximity, and permits covalent bonding of their side chains. A disulfide bond contributes to the stability of the three-dimensional shape of the protein molecule.

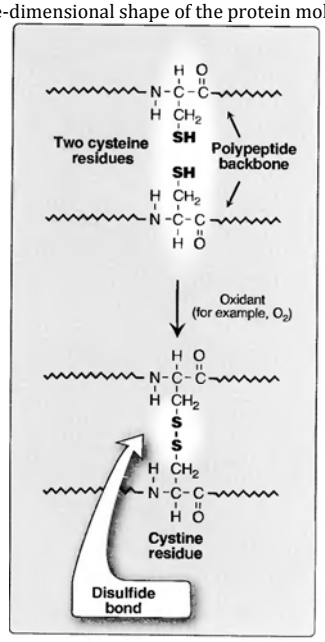


Figure 1
Formation of a disulfide bond by the oxidation of two cysteine residues, producing one cystine residue.

For example, many disulfide bonds are found in proteins such as immunoglobulins that are secreted by cells. [Note: These strong, covalent bonds help stabilize the structure of proteins, and prevent them from becoming denatured in the extracellular environment.]

2. Hydrophobic interactions: Amino acids with nonpolar side chains tend to be located in the interior of the polypeptide molecule, where they associate with other hydrophobic amino acids (Figure 2). In contrast, amino acids with polar or charged side chains tend to be located on the surface of the molecule in contact with the polar solvent. [Note: Proteins located in nonpolar (lipid) environments, such as a membrane, exhibit the reverse arrangement--that is, hydrophilic amino acid side chains are located in the interior of the polypeptide, whereas hydrophobic amino acids are located on the surface of the molecule in contact with the non-polar environment. In each case, the segregation of R-groups occurs that is energetically most favorable.

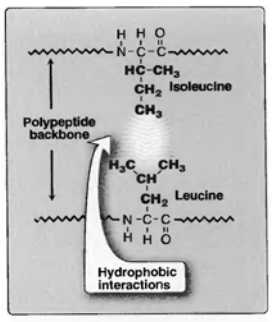


Figure 2
Hydrophobic interactions between amino acids with nonpolar side chains.

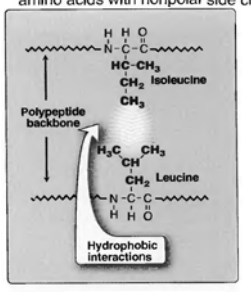


Figure 3 Hydrophobic interactions between amino acids with nonpolar side chains

3. Hydrogen bonds: Amino acid side chains containing oxygen- or nitrogen-bound hydrogen, such as in the alcohol groups of serine and threonine, can form hydrogen bonds

with electron-rich atoms, such as the oxygen of a carboxyl group or carbonyl group of a peptide bond. Formation of hydrogen bonds between polar groups on the surface of proteins and the aqueous solvent enhances the solubility of the protein(Figure 3).

4. Ionic interactions: Negatively charged groups, such as the carboxyl group ($-COO^-$) in the side chain of aspartate or glutamate, can interact with positively charged groups, such as the amino group ($-NH_3^+$) in the side chain of lysine.

Protein Folding: Interactions between the side chains of amino acids determine how a long polypeptide chain folds into the intricate three-dimensional shape of the functional protein. Protein folding, which occurs within the cell in seconds to minutes, employs a shortcut through the maze of all folding possibilities. As a peptide folds, its amino acid side chains are attracted and repulsed according to their chemical properties. For example, positively and negatively charged side chains attract each other. Conversely, similarly charged side chains repel each other. In addition, interactions involving hydrogen bonds, hydrophobic interactions, and disulfide bonds all seek to exert an influence on the folding process. This process of trial and error tests many, but not all, possible configurations, seeking a compromise in which attractions outweigh repulsions. This results in a correctly folded protein with a low energy state.

Role of Chaperons in Protein Folding:

It is generally accepted that the information needed for correct protein folding is contained in the primary structure of the polypeptide. Given that premise, it is difficult to explain why most proteins when denatured do not resume their native conformations under favorable environmental conditions. One answer to this problem is that a protein begins to fold in stages during its synthesis, rather than waiting for synthesis of the entire chain to be totally completed. This limits competing folding configurations made available by longer stretches of nascent peptide. In addition, a specialized group of proteins, named "chaperones," are required for the proper folding of many species of proteins. The chaperones--also known as "heat shock" proteins--interact with the polypeptide at various stages during the folding process. Some chaperones are important in keeping the protein unfolded until its synthesis is finished, or act as catalysts by increasing the rates of the final stages in the folding process. Others protect proteins as they fold so that their vulnerable, exposed regions do not become tangled in unproductive encounters.

Quaternary Structure of Protein:

Many proteins consist of a single polypeptide chain, and are defined as monomeriў proteins. However, others may consist of two or more polypeptide chains that may be structurally identical or totally unrelated. The arrangement of these polypeptide subunits is called the quaternary structure of the protein. [Note: If there are two subunits, the protein is called "dimeric", if three subunits "trimeric", and, if several subunits, "multimeric."] Subunits are held together by noncovalent interactions (for example, hydrogen bonds, ionic bonds, and hydrophobic interactions). Subunits may either function independently of each other, or may work cooperatively, as in hemoglobin, in which the binding of oxygen to one subunit of the tetramer increases the affinity of the other subunits for oxygen.

Denaturation of Protein:

Protein denaturation results in the unfolding and disorganization of the protein's secondary and tertiary structures, which are not accompanied by hydrolysis of peptide bonds. Denaturing agents include heat, organic solvents, mechanical mixing, strong acids or bases, detergents, and ions of heavy metals such as lead and mercury. Denaturation may, under ideal conditions, be reversible, in which case the protein refolds into its original native structure when the denaturing agent is removed. However, most proteins, once denatured, remain permanently disordered. Denatured proteins are often insoluble and, therefore, precipitate from solution.

I2.STABILITY OF NUCLIEC ACIDS:

In the Watson-Crick structure, the two chains or strands of the helix are antiparallel; their 5',3'-phosphodiester bonds run in opposite directions. Later work with DNA polymerases provided experimental evidence, confirmed by x-ray crystallography, that the strands are indeed antiparallel.

To account for the periodicities observed in the x-ray diffraction pattern, Watson and Crick used molecular models to show that the vertically stacked bases inside the double helix would be 0.34 nm apart and that the secondary repeat distance of about 3.4 nm could be accounted for by the presence of 10 (now 10.5) nucleotide residues in each complete turn of the double helix . The two antiparallel polynucleotide chains of double-helical DNA are not identical in either base sequence or composition. Instead they are complementary to each other. Wherever adenine appears in one chain, thymine is found in the other; similarly, wherever guanine is found in one chain, cytosine is found in the other.

The DNA double helix or duplex is held together by two sets of forces, as described earlier: hydrogen bonding between complementary **base pairs and base-stacking interactions**.

The specificity that maintains a given base sequence in each DNA strand is contributed entirely by the hydrogen bonding between base pairs. The basestacking interactions, which are largely nonspecific with respect to the identity of the stacked bases, make the major contribution to the stability of the double helix.

J1. METABOLISM OF CARBOHYDRARES: CITRIC ACID CYCLE

The citric acid cycle (Krebs cycle or tricarboxylic acid-TCA cycle) is the most important metabolic pathway for the energy supply to the body. About 65-70% of the ATP is synthesized in Krebs cycle. **Citric acid cycle essentially involves the oxidation of acetyl CoA to CO_2 and H_2O**. This cycle utilizes about two-thirds of total oxygen consumed by the body. The name TCA cycle is used, since, at the outset of the cycle, tricarboxylic acids (citrate, cisaconitate and isocitrate) participate.

TCA cycle-the central metabolic pathway

The citric acid cycle is the final common oxidative pathway for carbohydrates, fats and amino acids. This cycle not only supplies energy but also provides many intermediates required for the synthesis of amino acids, glucose, heme etc.

Krebs cycle is the most important central pathway connecting almost all the individual metabolic pathways (either directly or indirectly).

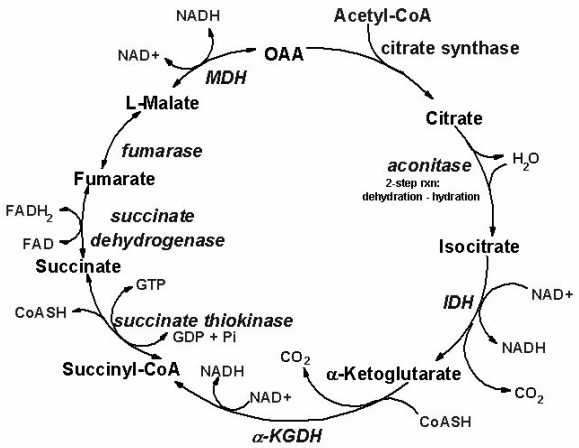
Brief history: The citric acid cycle was proposed by Hans Adolf Krebs in 1937, based on the studies of oxygen consumption in pigeon breast muscle. The cycle is named in his honour and he was awarded the Nobel Prize for Physiology and Medicine in 1953.

Location of TCA cycle

The enzymes of TCA cycle are located in **mitochondrial matrix**, in close proximity to the electron transport chain. This enables the synthesis of ATP by oxidative phosphorylation without any hindrance.

TCA cycle-an overview

Krebs cycle basically involves the combination of a two carbon acetyl CoA with a four carbon oxaloacetate to produce a six carbon tricarboxylic acid, citrate. In the reactions that follow, the two carbons are oxidized to CO_2 and oxaloacetate is regenerated and recycled. **Oxaloacetate** is considered to play a catalytic role in citric acid cycle.



TCA cycle-an open cycle

Krebs cycle is a cyclic process. However, it should not be viewed as a closed circle, since many compounds enter the cycle and leave. TCA cycle is comparable to a heavy traffic circle in a national highway with many connecting roads. Each intermediate of the cycle connecting another pathway is a road!

Reactions of citric acid cycle

Oxidative decarboxylation of pyruvate to acetyl CoA by pyruvate dehydrogenase complex is discussed above. This step is a connecting link between glycolysis and TCA cycle. A few authors, however, describe the conversion of pyruvate to acetyl CoA along with citric acid cycle. The events of TCA cycle are described hereunder.

Step 1: Formation of citrate: Krebs cycle proper starts with the condensation of acetyl CoA and oxaloacetate, catalysed by the enzyme citrate synthase. This is an aldol condensation reaction leading to the formation of an intermediate citryl CoA, which, on hydrolysis, yields citrate. Citrate is freely permeable across the mitochondrial membrane. It serves as a good source of cytosolic acetyl CoA which is used for the synthesis of fatty acids. Citrate inhibits phosphofructokinase and activates acetyl CoA carboxylase. The latter is the key enzyme in fatty acid synthesis.

Step 2 and 3: Citrate is isomerized to isocitrate by the enzyme aconitase. This is achieved in a two stage reaction of dehydration followed by hydration through the formation of an intermediate-cisaconitate.

Step 4 and 5: **Formation of α-ketoglutarate**: The enzyme isocitrate dehydrogenase (lCD) catalyses the conversion (oxidative decarboxylation) of isocitrate to oxalosuccinate and then to aketoglutarate. The formation of NADH and the liberation of C02 occurs at this stage. Three types of ICD are known. (1): NAD^+ specific, located only in mitochondria. (2): $NADP^+$ specific present in mitochondria, and cytosol (3). The mitochondrial NAD^+ dependent ICD is responsible for the conversion of isocitrate to α-ketoglutarate.

Step 6: Conversion of α-ketoglutarate to succinyl CoA occurs through oxidative decarboxylation, catalysed by a-ketoglutarate dehydrogenase complex. This enzyme is dependent on five cofactors- TPP, lipoamide, NAD+, FAD and CoA. The mechanism of the reaction is analogous to the conversion of pyruvate to acetyl CoA. At this stage of the TCA cycle, second NADH is produced and the second CO_2 is liberated.

Step 7: Formation of succinate: Succinyl CoA is converted to succinate by succinate thiokinase. This reaction is coupled with the phosphorylation of GDP to GTP. This is a substrate level phosphorylation. GTP is converted to ATP by the enzyme nucleoside diphosphate kinase.

GTP + ADP ⟷ ATP + GDP

Step 8: **Conversion of succinate to fumarate** : Succinate is oxidized by succinate dehydrogenase to fumarate. This reaction results in the production of FADH2 and not NADH. This is due to the fact that the reducing power of succinate is not adequate to reduce NAD+, hence FAD is utilized.

Step 9: Formation of malate: The enzyme fumarase catalyses the conversion of fumarate to malate with the addition of H_2O.

Step 10: Conversion of malate to oxaloacetate: Malate is then oxidized to oxaloacetate by malate dehydrogenase. The third and final synthesis of NADH occurs at this stage. The oxaloacetate is regenerated which can combine with another molecule of acetyl CoA and continue the cycle.

Summary of TCA cycle: The events of Krebs cycle may be summarized as

Acetyl CoA + 3 NAD^+ + FAD + GDP + Pi + $2H_2O$→ $2CO_2$ + 3NADH + $3H^+$ + $FADH_2$ + GTP + CoA

Requirement of O_2 by TCA cycle: There is no direct participation of oxygen in Krebs cycle. However, the cycle operates only under aerobic conditions. This is due to the fact that NAD^+ and FAD (from NADH and $FADH_2$ respectively) required for the operation of the cycle can be regenerated in the respiratory chain only in the presence of O_2. Therefore, citric acid cycle is strictly aerobic in contrast to glycolysis which operates in both aerobic and anaerobic conditions.

Energetics of citric acid cycle: During the process of oxidation of acetyl CoA via citric acid cycle, 4 reducing equivalents (3 as NADH and one as $FADH_2$) are produced. Oxidation of 1 NADH by electron transport chain coupled with oxidative phosphorylation results in the synthesis of 3 ATP, whereas $FADH_2$ leads to the formation of 2 ATP. Besides, there is one substrate level phosphorylation. **A total of twelve ATP are produced from one acetyl CoA**.

Inhibitors of Krebs cycle

The important enzymes of TCA cycle inhibited by the respective inhibitors are listed

Enzyme	Inhibitor
Aconitase	Fluoroacetate (non-competitive)
α.-Ketoglutarate dehydrogenase	Arsenite (non-competitive)
Succinate dehydrogenase	Malonate (competitive)

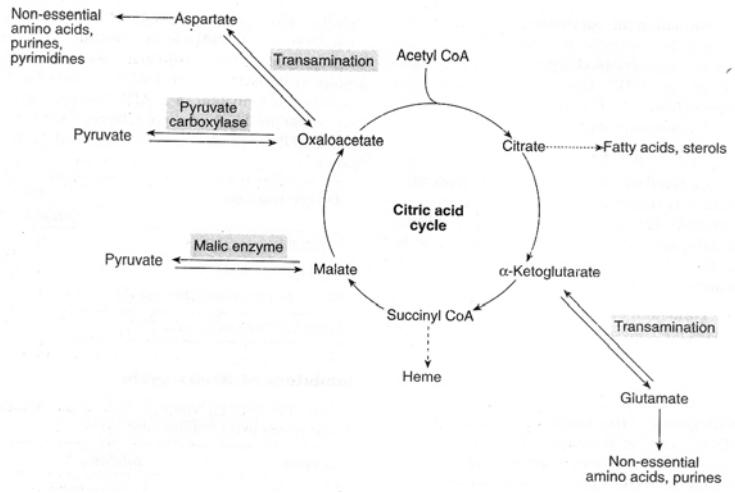
Regulation of citric acid cycle

The cellular demands of ATP are crucial in controlling the rate of citric acid cycle. The regulation is brought about either by enzymes or the levels of ADP. Three enzymes-namely **citrate synthase, isocitrate dehydrogenase and α-ketoglutarate dehydrogenase**-regulate citric acid cycle.

1. **Citrate synthase** is inhibited by ATP, NADH, acyl CoA and succinyl CoA.
2. **Isocitrate dehydrogenase** is activated by ADP and inhibited by ATP and NADH.
3. **α-Ketoglutarate dehydrogenase** is inhibited by succinyl CoA and NADH.
4. **Availability of ADP** is very important for the citric acid cycle to proceed. This is due to the fact that unless sufficient levels of ADP are available, oxidation (coupled with phosphorylation of ADP to ATP) of NADH and FADH2 through electron transport chain stops. The accumulation of NADH and FADH2 will lead to inhibition of the enzymes (as stated above) and also limits the supply of NAD^+ and FAD which are essential for TCA cycle to proceed.

Amphibolic nature of the citric acid cycle

The citric acid cycle is the final common oxidative pathway in the living cells. The cycle also provides various intermediates for the synthesis of many compounds needed by the body. Krebs cycle is **both catabolic and anabolic in nature**, hence regarded as amphibolic.



The most important synthetic (anabolic) reactions connected with TCA cycle are given

1. Oxaloacetate and α-ketoglutarate, respectively, serve as precursors for the synthesis of' aspartate and glutamate which, in turn, are required for the synthesis of other non-essential amino acids, purines and pyrimidines.
2. Succinyl CoA is used for the synthesis of porphyrins and heme.
3. Mitochondrial citrate is transported to the cytosol, where it is cleaved to provide acetyl CoA for the biosynthesis of fatty acids, sterols etc.

Energetics of glucose oxidation: When a molecule of glucose (6 carbon) undergoes glycolysis, 2 molecules of pyruvate or lactate (3 carbon) are produced. Pyruvate is oxidatively decarboxylated to acetyl CoA (2 carbon) which enters the citric acid cycle and gets completely oxidized to CO_2 and H_2O. The overall process of glucose being completely oxidized to CO_2 and H_2O via glycolysis and citric acid cycle is

$C_6H_{12}O_6 + 6O_2 + 38ADP + 38Pi \rightarrow 6CO_2 + 6H_2O + 38ATP$

When a molecule of glucose is burnt in a calorimeter, 2,780 KJ of heat is liberated. In the living system, energy is trapped leading to the synthesis of 38 ATP which is equivalent to 1,159 KJ (1 ATP has high energy bond equivalent to 30.5 KJ). That is, about 48% of the energy in glucose combustion is actually captured for ATP generation.

J2. GLUCONEOGENESIS

The synthesis of glucose or glycogen from non-carbohydrate compounds is known as gluconeogenesis. The major substrates/precursors for gluconeogenesis are lactate, pyruvate, glucogenic amino acids, propionate and glycerol.

Location of gluconeogenesis: Gluconeogenesis occurs mainly in the cytosol, although some precursors are produced in the mitochondria. Gluconeogenesis mostly takes place in liver and, to some extent, in kidney matrix (about one-tenth of liver capacity).

Importance of gluconeogenesis: Glucose occupies a key position in the metabolism and its continuous supply is absolutely essential to the body for a variety of functions

1. Brain and central nervous system, erythrocytes, testes and kidney medulla are dependent on glucose for continuous supply of energy. Human brain alone requires about 120 g of glucose per day, out of about 160 g needed by the entire body.
2. Glucose is the only source that supplies energy to the skeletal muscle, under anaerobic conditions.
3. All other carbohydrates are synthesized from glucose, e.g. lactose, amino sugars etc.
4. Glucose is converted to glycerol and utilized for the synthesis of fat.
5. In fasting even more than a day, gluconeogenesis must occur to meet the basal requirements of the body for glucose and to maintain the intermediates of citric acid cycle. This is essential for the survival of humans and other animals.
6. Certain metabolites produced in the tissues accumulate in the blood, e.g. lactate, glycerol, propionate etc. Gluconeogenesis effectively clears them from the blood.

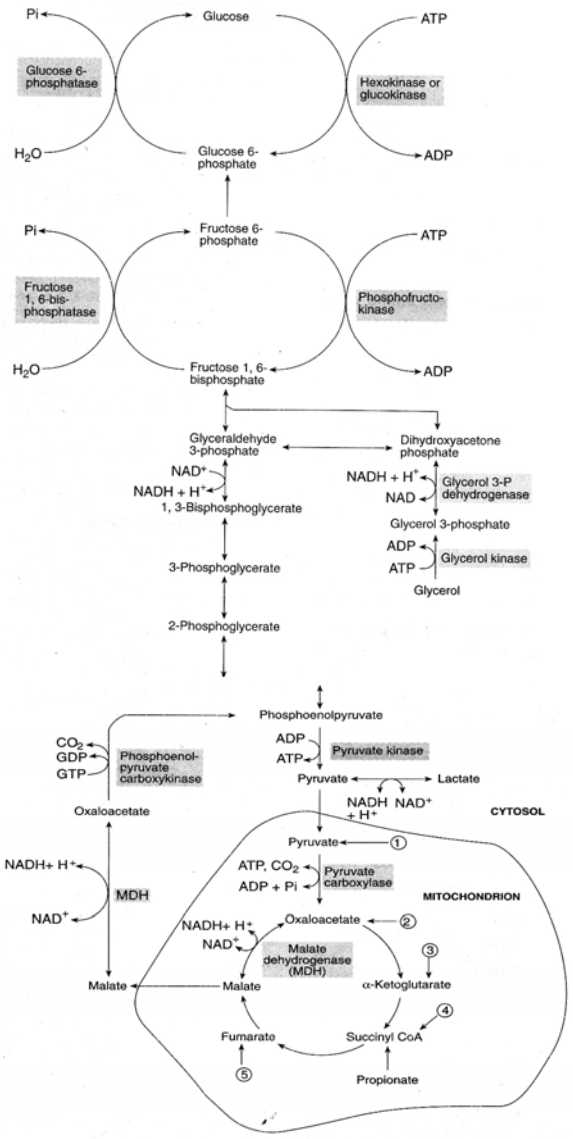
Reactions of gluconeogenesis: Gluconeogenesis closely resembles the reversed pathway of glycolysis, although it is not the complete reversal of glycolysis. Essentially, 3 (out of 10) reactions of glycolysis are irreversible. The seven reactions are common for both glycolysis and gluconeogenesis (**Fig.**). The three irreversible steps of glycolysis are catalysed by the enzymes, namely hexokinase, phosphofructokinase and pyruvate kinase. These three stages-by passed by alternate enzymes specific to gluconeogenesis are dicussed.

1. Conversion of pyruvate to phosphoenolpyruvate: This takes place in two steps. Pyruvate carboxylase is a biotin-dependent mitochondrial enzyme that converts pyruvate to oxaloacetate in presence of ATP and CO_2. This enzyme regulates gluconeogenesis and requires acetyl CoA for its activity.

Oxaloacetate is synthesized in the mitochondrial matrix. It has to be transported to the cytosol to be used in gluconeogenesis, where the rest of the pathway occurs. Due to membrane impermeability, oxaloacetate cannot diffuse out of the mitochondria. It is converted to malate and then transported to the cytosol. Within the cytosol, oxaloacetate is regenerated. The reversible conversion of oxaloacetate and malate is catalysed by malate dehydrogenase, an enzyme present in both mitochondria and cytosol.

In the cytosol, phosphoenolpyruvate carboxykinase converts oxaloacetate to phosphoenolpyruvate. GTP or ITP (not ATP) is used in this reaction and the CO_2 (fixed by carboxylase) is liberated. For the conversion of pyruvate to phosphoenol pyruvate, 2 ATP equivalents are utilized. This is in contrast to only one ATP that is liberated in glycolysis for this reaction.

2. Conversion of fructose 1, 6-bisphosphate to fructose 6-phosphate: Phosphoenolpyruvate undergoes the reversal of glycolysis until fructose 1, 6-bisphosphate is produced. The enzyme fructose 1, 6-bisphosphatase converts fructose 1, 6bisphosphate to fructose 6-phosphate. This enzyme requires $Mg2^+$ ions. Fructose 1, 6-bisphosphatase is absent in smooth muscle and heart muscle. This enzyme is also regulatory in gluconeogenesis.



3. Conversion of glucose 6-phosphate to glucose: Glucose 6-phosphatase catalyses the conversion of glucose 6-phosphate to glucose. The presence or absence of this enzyme in a tissue determines whether the tissue is capable of contributing glucose to the blood or not. It is mostly present in liver and kidney but absent in muscle, brain and adipose tissue.

The overall summary of gluconeogenesis for the conversion of pyruvate to glucose is shown:

2 Pyruvate + 4ATP + 2GTP + 2NADH+ $2H^+$ +$6H_2O$ → Glucose + $2NAD^+$ + 4ADP + 2GDP + 6Pi + $6H^+$

After the formation of glucose, its conversion to glycogen occurs through the formation of uridine diphosphate glucose (UDP glucose) by the enzyme glycogen synthetase.

Gluconeogenesis from amino acids

The carbon skeleton of glucogenic amino acids (all except leucine and lysine) results in the formation of pyruvate or the intermediates of citric acid cycle which, ultimately, result in the synthesis of glucose.

Gluconeogenesis from glycerol

Glycerol is liberated mostly in the adipose tissue by the hydrolysis of fats (triacylglycerols). The enzyme glycerokinase (found in liver and kidney, absent in adipose tissue) activates glycerol to glycerol 3-phosphate. The latter is converted to dihydroxyacetone phosphate by glycerol 3-phosphate dehydrogenase. Dihydroxyacetone phosphate is an intermediate in glycolysis which can be conveniently used for glucose production.

Gluconeogenesis from propionate

Oxidation of odd chain fatty acids and the breakdown of some amino acids (methionine, isoleucine) yields a three carbon propionyl CoA. Propionyl CoA carboxylase acts on this in presence of ATP and biotin and converts to methyl malonyl CoA which is then converted to succinyl CoA in presence of B12 coenzyme. Succinyl CoA formed from propionyl CoA enters gluconeogenesis via citric acid cycle. Propionate is an important precursor for gluconeogenesis in ruminant animals such as cattle. Propionate is activated to propionyl CoA which is then converted to glucose as described above.

Gluconeogenesis from lactate (Cori cycle)

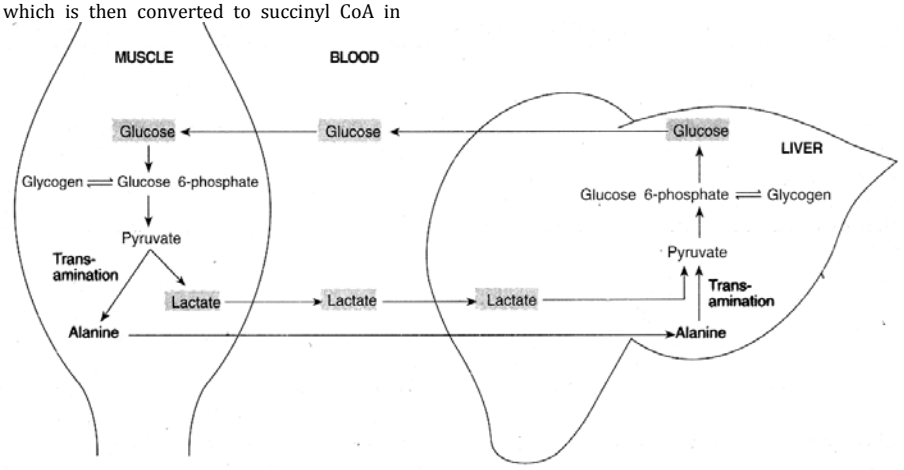
Lactate produced by active skeletal muscle is a major precursor for gluconeogenesis. .Under anaerobic conditions, pyruvate is reduced to lactate by lactate dehydrogenase (LDH)

LDH

Pyruvate + NADH + H^+ → Lactate + NAD^+

Lactate is a dead end in glycolysis, since it must be reconverted to pyruvate for its further metabolism. The very purpose of lactate production is to regenerate NADH so that glycolysis proceeds uninterrupted in skeletal muscle. Lactate or pyruvate produced in the muscle cannot be utilized for the synthesis of glucose due to the absence of the key enzymes of gluconeogenesis (glucose 6-phosphatase and fructose 1, 6-bisphosphatase).

The plasma membrane is freely permeable to lactate. Lactate is carried from the skeletal muscle through blood and handed over to liver, where it is oxidized to pyruvate. Pyruvate, so produced, is converted to glucose by gluconeogenesis, which is then transported to the skeletal muscle.



The cycle involving the synthesis of glucose in liver from the skeletal muscle lactate and the reuse of glucose thus synthesized by the muscle for energy purpose is known as Cori cycle (**Fig**).

Glucose-alanine cycle:

There is a continuous transport of amino acids from muscle to liver, which predominantly occurs during starvation. Alanine dominates among the amino acids. It is postulated that. pyruvate in skeletal muscle undergoes transamination to produce alanine. Alanine is transported to liver and used for gluconeogenesis. This cycle is referred to as glucose-alanine cycle.

Regulation of gluconeogenesis: The hormone glucagon and the availability of substrates mainly regulate gluconeogenesis, as discussed hereunder.

1. Influence of glucagon: This is a hormone, secreted by a-cells of the pancreatic islets. Glucagon stimulates gluconeogenesis by two mechanisms

a. Active form of pyruvate kinase is converted to inactive form through the mediation of cyclic AMP, brought about by glucagon. Decreased pyruvate kinase results in the reduced conversion of phosphoenol pyruvate to pyruvate and the former is diverted for the synthesis of glucose.

b. Glucagon reduces the concentration of fructose 2, 6-bisphosphate. This compound allosterically inhibits phosphofructokinase and activates fructose 1, 6-bisphosphatase, both favour increased gluconeogenesis.

Generally, when gluconeogenesis is more active, glycolysis is very low. This is a good example of cellular economy.

2. Availability of substrates: Among the various substrates, glucogenic amino acids have stimulating influence on gluconeogenesis. This is particularly important in a condition like diabetes mellitus (decreased insulin level) where amino acids are mobilized from muscle protein for the purpose of gluconeogenesis.

3. Acetyl CoA promotes gluconeogenesis: During starvation-due to excessive lipolysis in adipose tissue--acetyl CoA accumulates in the liver. Acetyl CoA allosterically activates pyruvate carboxylase resulting in enhanced glucose production.

Gluconeogenesis from fat? It is often stated that glucose cannot be synthesized from fat. In a sense, it is certainly true, since the fatty acids (most of them being even chain), on oxidation, produce acetyl CoA which cannot be converted to pyruvate. Further, the two carbons of acetyl CoA disappear as 2 moles of CO_2 in TCA cycle. Therefore, even chain fatty acids cannot serve as precursors for glucose formation.

However, the glycerol released from lipolysis and the propionate obtained from the oxidation of odd chain fatty acids are good substrates for gluconeogenesis, as discussed above.

J3. GLYCOGEN METABOLISM

Glycogen is the storage form of glucose in animals, as is starch in plants. It is stored mostly in liver (6-8%) and muscle (1-2%). Due to more muscle mass, the quantity of glycogen in muscle (250 g) is about three times higher than that in the liver (75 g).

Functions of glycogen

The prime function of liver glycogen is to maintain the blood glucose levels, particularly between meals. Liver glycogen stores increase in a well-fed state which are depleted during fasting. Muscle glycogen serves as a fuel reserve for the supply of ATP during muscle contraction.

Structure of glycogen-in brief

Glycogen is a homopolysaccharide composed of α-D-glucose (up to 100,000 residues may be present). The glucose units are held together by α-1, 4 linkages. After 8-10 residues of glucose, a branch is formed with α-1, 6 glycosidic linkage. Glycogen is stored as granules in the cytosol, where most of the enzymes of glycogen synthesis and breakdown are present.

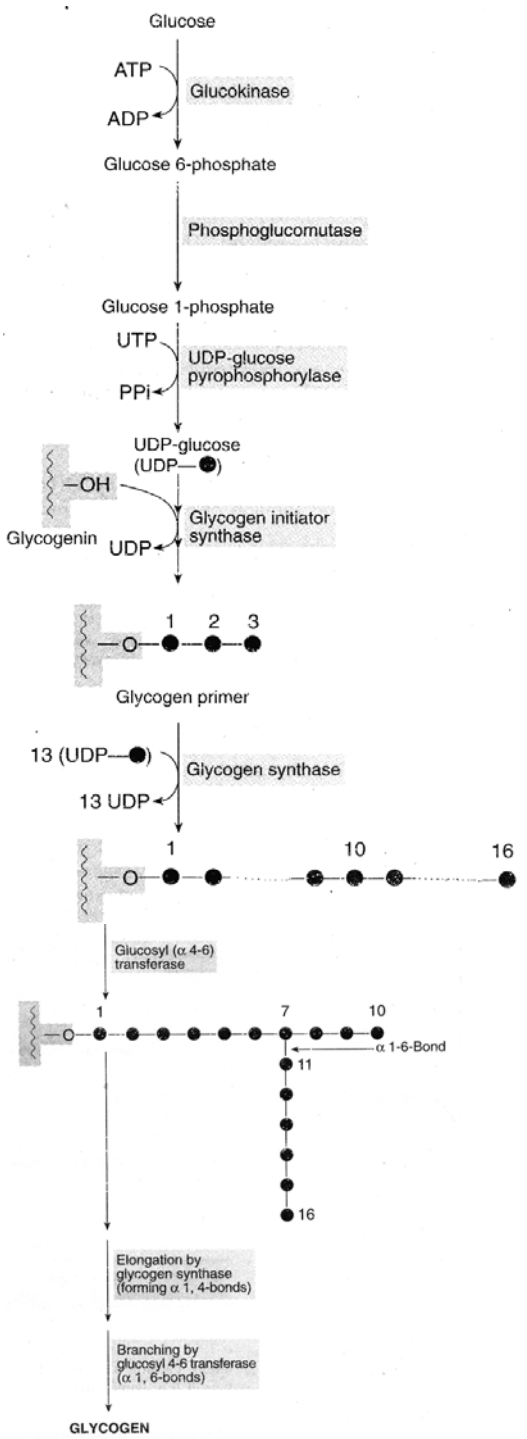
GLYCOGENESIS

The synthesis of glycogen from glucose is glycogenesis (**Fig**.). Glycogenesis takes place in the cytosol and requires ATP and UTP, besides glucose.

1. Synthesis of UDP-glucose: The enzymes hexokinase (in muscle) and glucokinase (in liver) convert glucose to glucose 6-phosphate. Phosphoglucomutase catalyses the conversion of glucose 6phosphate to glucose 1-phosphate. Uridine diphosphate glucose (UDPG) is synthesized from glucose 1-phosphate and UTP by UDP-glucose pyrophosphorylase. Pyrophosphate (PPi) produced in this reaction is hydrolysed to inorganic phosphate (Pi) by pyrophosphatase. This will ensure the optimal synthesis of UDPG.

2. Requirement of primer to initiate glycogenesis: A small fragment of pre-existing glycogen must act as a '**primer'** to initiate glycogen synthesis. It is recently found that in the absence of glycogen primer, a specific protein-namely '**glycogenin**' can accept glucose from UDPG. The hydroxyl group of the amino acid tyrosine of glycogenin is the site at which the initial glucose unit is attached. The enzyme glycogen initiator synthase transfers the first molecule of glucose to glycogenin. Then glycogenin itself takes up a few glucose residues to form a fragment of primer which serves as an acceptor for the rest of the glucose molecules.

3. Glycogen synthesis by glycogen synthase: Glycogen synthase is responsible for the formation of 1, 4-glycosidic linkages. This enzyme transfers the glucose from UDP-glucose to the non-reducing end of glycogen to form α-1, 4 linkages. The UDP released can be converted back to UTP by nucleoside diphosphate kinase.



4. Formation of branches in glycogen: Glycogen synthase can catalyse the synthesis of a linear unbranched molecule with 1, 4 -α-glycosidic linkages. Glycogen, however, is a branched treelike structure. The formation of branches is brought about by the action of a branching enzyme, namely glucosyl α-4-6 transferase (amylo α-1, 4 → 1, 6 transglucosidase). This enzyme transfers a small fragment of five to eight glucose residues from the non-reducing end of glycogen chain (by breaking α-1, 4 linkages) to another glucose residue where it is linked by α-1, 6 bond. This leads to the formation of a new non-reducing end, besides the existing one. Glycogen is further elongated and branched, respectively, by the enzymes glycogen synthase and glucosyl 4-6 transferase.

The overall reaction of the glycogen synthesis for the addition of each glucose residue is

$(Glucose)_n$ + Glucose + 2ATP → $(Glucose)_{n+l}$ + 2 ADP + P_i

Of the two ATP utilized, one is required for the phosphorylation of glucose while the other is needed for conversion of UDP to UTP.

GLYCOGENOLYSIS

The degradation of stored glycogen in liver and muscle constitutes glycogenolysis. The pathway for the synthesis and degradation of glycogen are not reversible. An independent set of enzymes present in the cytosol carry out glycogenolysis. Glycogen is degraded by breaking α-1, 4- and α-1, 6-glycosidic bonds (**Fig**).

1. Action of glycogen phosphorylase: The α- 1, 4 glycosidic bonds (from the non-reducing ends) are cleaved sequentially by the enzyme glycogen phosphorylase to yield glucose l-phosphate. This process-called phosphorolysis-continues until four glucose residues remain on either side of branching point (α-l, 6-glycosidic link). The glycogen so formed is known as limit dextrin which cannot be further degraded by phosphorylase. Glycogen phosphorylase possesses a molecule of pyridoxal phosphate, covalently bound to the enzyme.

2. Action of debranching enzyme: The branches of glycogen are cleaved by two enzyme activities present on a single polypeptide called **debranching enzyme**, hence it is a **bifunctional enzyme**.

Glycosyl 4:4 transferase (oligo α-1, 4 → 1, 4 glucan transferase) activity removes a fragment of three or four glucose residues attached at a branch and transfers them to another chain. Here, one α-l, 4-bond is cleaved and the same a- 1, 4 bond is made, but the places are different.

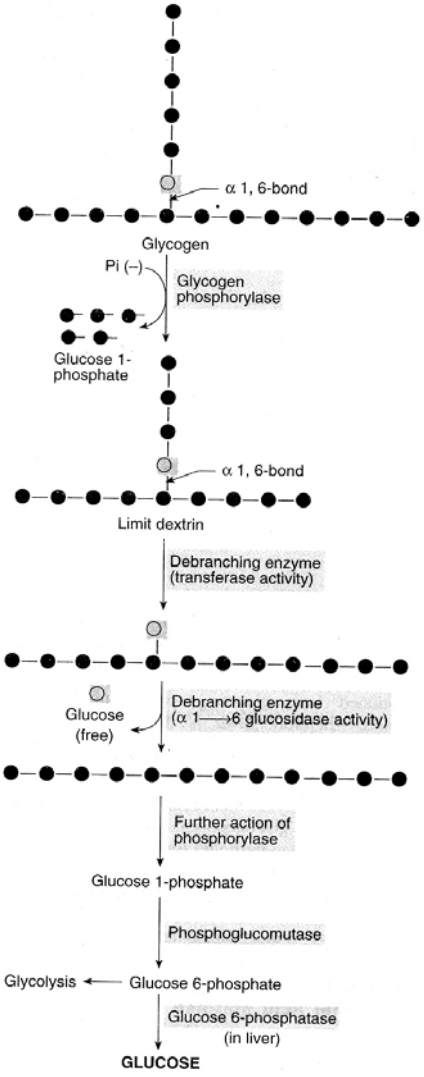
Amylo α-l, 6-glucosidase breaks the α-l, 6 bond at the branch with a single glucose residue and **releases a free glucose**.

The remaining molecule of glycogen is again available for the action of phosphorylase and debranching enzyme to' repeat the reactions stated in 1 and 2.

3. Formation of glucose 6-phosphate and glucose: Through the combined action of glycogen phosphorylase and debranching enzyme, glucose l-phosphate and free glucose in ratio of 8: 1 are produced. Glucose l-phosphate is converted to glucose 6-phosphate by the enzyme phospho-glucomutase.

The fate of glucose 6-phosphate depends on the tissue. The liver, kidney and intestine contain the enzyme **glucose 6-phosphatase** that cleaves glucose 6-phosphate to glucose. This enzyme is absent in muscle and brain, hence free glucose cannot be produced from glucose 6-phosphate in these tissues. Therefore, liver is the major glycogen storage organ to provide glucose into the circulation to be utilised by various tissues.

In the peripheral tissues, glucose 6-phosphate produced by glycogenolysis will be used for glycolysis. It may be noted that though glucose-6-phosphatase is absent in muscle, some amount of free glucose (8-10% of glycogen) is produced in glycogenolysis due to the action of debranching enzyme (α-l, 6-glucosidase activity).



Degradation of glycogen by lysosomal acid maltase

Acid maltase or α-l, 4-glucosidase is a lysosomal enzyme. This enzyme continuously degrades a small quantity of glycogen. The significance of this pathway is not very clear. However, it has been observed that the deficiency of lysosomal enzyme α-l, 4 glucosidase results in glycogen accumulation, causing a serious glycogen storage disease type II (i.e. Pompe's disease).

Regulation of glycogenesis and glycogenolysis

A good coordination and regulation of glycogen synthesis and its degradation is essential to maintain the blood glucose levels. Glycogenesis and glycogenolysis are, respectively, controlled by the enzymes glycogen synthase

and glycogen phosphorylase. Regulation of these enzymes is accomplished by three mechanisms

1. Allosteric regulation of glycogen metabolism: There are some metabolites that allosterically regulate the activities of glycogen synthase and glycogen phosphorylase. The control is carried out in such a way that glycogen synthes!s is increased when substrate availability and energy levels are high. On the other hand, glycogen breakdown is enhanced when glucose concentration and energy levels are low. The allosteric regulation of glycogen metabolism is depicted in Fig. In a well-fed state, the availability of glucose 6-phosphate is high which allosterically activates glycogen synthase for more glycogen synthesis. On the other hand, glucose 6-phosphate and ATP allosterically inhibit glycogen phosphorylase. Free glucose in liver also acts as an allosteric inhibitor of glycogen phosphorylase.

2. Hormonal regulation of glycogen metabolism: The hormones, through a complex series of reactions, bring about covalent modification, namely phosphorylation and dephosphorylation of enzyme proteins which, ultimately control glycogen synthesis or its degradation.

cAMP as second messenger for hormones: The hormones like epinephrine and norepinephrine, and glucagon (in liver) activate adenylate cyclase to increase the production of cAMP. The enzyme phosphodiesterase breaks down cAMP. The hormone insulin increases the phosphodiesterase activity in liver and lowers the cAMP levels.

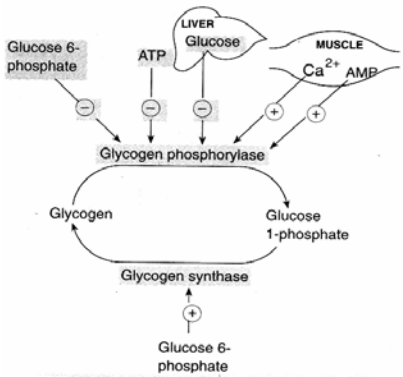


Fig.: Allosteric regulation of glycogenolysis and glycogenesis (⊖ : Inhibition; ⊕ : Activation).

Regulation of glycogen synthesis by cAMP: The glycogenesis is regulated by glycogen synthase. This enzyme exists in two forms-**glycogen synthase 'a'**-which is not phosphorylated and most active, and secondly, **glycogen synthase 'b'** as phosphorylated inactive form. Glycogen synthase 'a' can be converted to 'b' form (inactive) by phsophorylation. The degree of phosphorylation is proportional to the inactive state of enzyme. The process of phosphorylation is catalysed by a cAMP dependent protein kinase. The protein kinase phosphorylates and inactivates glycogen synthase by converting 'a' form to 'b' form. The glycogen synthase '**b**' can be converted back to synthase '**a**' by protein phosphatase I.

In Fig. the inhibition of glycogen synthesis brought by epinephrine (also norepinephrine) and glucagon through cAMP by converting active glycogen synthase 'a' to inactive synthase 'b'.

Regulation of glycogen degradation by cAMP: The hormones like epinephrine and glucagon bring about glycogenolysis by their action on glycogen phosphorylase through cAMP. Glycogen phosphorylase exists in two forms, an active 'a' form and inactive form 'b'.

The cAMP-formed due to hormonal stimulusactivates cAMP dependent protein kinase. This active protein kinase phosphorylates inactive form of glycogen phsophorylase kinase to active form. (The enzyme protein phosphatase removes phosphate and inactivates phosphorylase kinase). The active phosphorylase kinase phosphorylates inactive glycogen phosphorylase 'b' to active glycogen phosphorylase 'a' which degrades glycogen. The enzyme protein phosphatase I can dephosphorylate and convert active glycogen phosphorylase 'a' to inactive 'b' form.

3. Effect of $Ca2^{+}$ ions on glycogenolysis: When the muscle contracts, Ca^{2+} ions are released from the sarcoplasmic reticulum. Ca^{2+} binds to calmodulin-calcium modulating protein and directly activates phosphorylase kinase without the involvement of cAMP-dependent protein kinase.

The overall effect of hormones in glycogen metabolism is that an elevated glucagon or epinephrine level increases glycogen degradation whereas elevated insulin results in increased glycogen synthesis.

J4. HEXOSE MONOPHOSPHATE SHUNT (Interconversion of hexoses and pentoses)

Hexose monophosphate pathway or **HMP shunt** is also called pentose phosphate pathway or phosphogluconate pathway. **This is an alternative pathway to glycolysis and TCA cycle for the oxidation of glucose**. However, HMP shunt is more anabolic in nature, since it is concerned with the biosynthesis of NADPH and pentoses.

HMP shunt-a unique multifunctional pathway

The pathway starts with glucose 6-phosphate. As such, no ATP is directly utilized or produced in HMP pathway. It is a unique multifunctional pathway, since there are several interconvertible substances produced which may proceed in different directions in the metabolic reactions.

Location of the pathway

The enzymes of HMP shunt are located in the cytosol. The tissues such as liver, adipose tissue, .adrenal gland, erythrocytes, testes and lactating mammary gland, are highly active in HMP shunt. Most of these tissues are involved in the biosynthesis of fatty acids and steroids which are dependent on the supply of NADPH.

The sequence of reactions of HMP shunt is divided into two phases-oxidative and non-oxidative.

1. Oxidative phase: Glucose 6-phosphate dehydrogenase (G6PD) is an NADP-dependent enzyme that converts glucose 6-phosphate to 6-phosphogluconolactone. The latter is then hydrolysed by the gluconolactone hydrolase to 6-phosphogluconate. The next reaction involving the synthesis of NADPH is catalysed by 6-phosphogluconate dehydrogenase to produce 3 keto 6-phosphogluconate which then undergoes decarboxylation to give ribulose 5-phosphate.

G6PD regulates HMP shunt: The first reaction catalysed by G6PD is most regulatory in HMP shunt. This enzyme catalyses an irreversible reaction. NADPH competitively inhibits G6PD. It is the ratio of NADPH/NAD^{+} that ultimately determines the flux of this cycle.

2. Non-oxidative phase: The non-oxidative reactions are concerned with the interconversion of three, four, five and seven carbon monosaccharides. Ribulose 5-phosphate is acted upon by an epimerase to produce xylulose 5-

phosphate while ribose 5-phosphate ketoisomerase converts ribulose 5-phosphate to ribose 5-phosphate.

The enzyme **transketolase** catalyses the transfer of two carbon moiety from xylulose 5-phosphate to ribose 5-phosphate to give a 3-carbon glyceraldehyde 3-phosphate and a 7 -carbon sedo heptulose 7-phosphate. Transketolase is dependent on the coenzyme thiamine pyrophosphate (TPP) and Mg2+ ions. Transaldolase brings about the transfer of a 3-carbon fragment (active dihydroxyacetone) from sedoheptu lose 7 -phosphate to glyceraldehyde 3-phosphate to give fructose 6-phosphate and four carbon erythrose 4-phosphate. Transketolase acts on xylulose 5-phosphate and transfers a 2-carbon fragment (glyceraldehyde) from it to erythrose 4-phosphate to generate fructose 6-phosphate and glyceraldehyde 3-phosphate.

Fructose 6-phosphate and glyceraldehyde 3-phosphate can be further catabolized through glycolysis and citric acid cycle. Glucose may also be synthesized from these two compounds.

An overview of HMP shunt is given in **Fig.** For the complete oxidation of glucose 6-phosphate to 6C02, we have to start with 6 molecules of glucose 6-phosphate. Of these 6, 5 moles are regenerated with the production of 12 NADPH.

The overall reaction may be represented as

6 Glucose 6-phosphate + 12 $NADP^+$ + 6H20 → 6C02 + 12 NADPH + 12H+ + 5 Glucose 6 phosphate.

Significance of HMP shunt:

HMP shunt is unique in generating two important products-pentoses and NADPH-needed for the biosynthetic reactions and other functions.

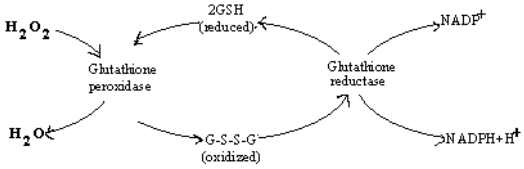
A. Importance of pentoses

In the HMP shunt, hexoses are converted into pentoses, the most important being ribose 5-phosphate. This pentose or its derivatives are useful for the **synthesis of nucleic acids** (RNA and DNA) and many **nucleotides** such as ATP, NAD^+, FAD and CoA.

Skeletal muscle is capable of synthesizing pentoses, although only the first few enzymes of HMP shunt are active. It, therefore, appears that the complete pathway of HMP shunt may not be required for the synthesis of pentoses.

B. Importance of NADPH

1. NADPH is required for the reductive **biosynthesis of fatty acids and steroids**, hence HMP shunt is more active in the tissues concerned with lipogenesis, e.g. adipose tissue, liver etc.
2. NADPH is used in the synthesis of certain amino acids involving the enzyme **glutamate dehydrogenase**.
3. There is a continuous production of H_2O_2 in the living cells which can chemically damage unsaturated lipids, proteins and DNA. This is, however, prevented to a large extent through **antioxidant reactions** involving NADPH. Glutathione mediated reduction of H_2O_2 is given hereunder



Glutathione (reduced, GSH) detoxifies H_2O_2, peroxidase catalyses this reaction. NADPH is responsible for the regeneration of reduced glutathione from the oxidized one.

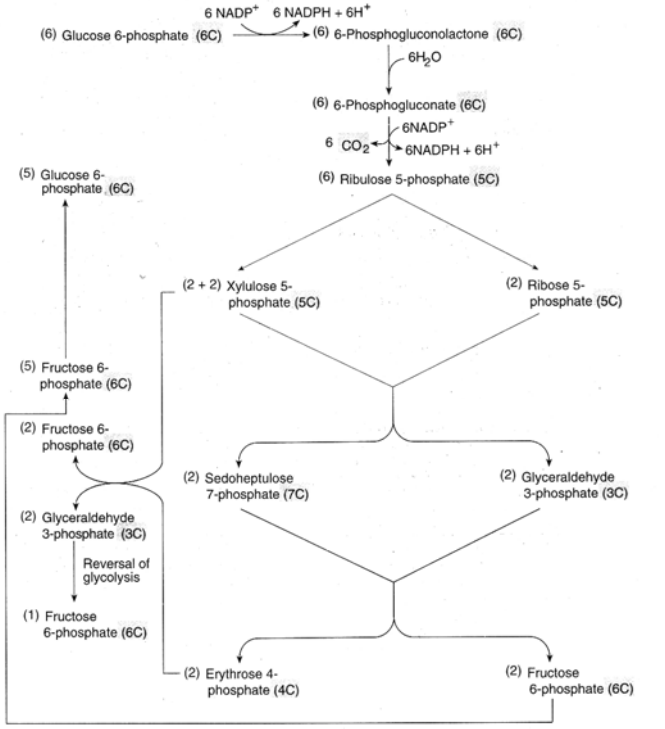


Fig. : Overview of hexose monophosphate shunt representing the number of molecules and the number of carbon aboms Note that of the 6-molecules of glucose 6-phosphate that enter HMP shunt, one molecule is oxidized as 5-molecules are finany recovered.

4. Microsomal cytochrome P_{450} system (in liver) brings about the **detoxification of drugs** and foreign compounds by hydroxylation reactions involving NADPH.
5. Phagocytosis is the engulfment of foreign particles, including microorganisms, carried out by white blood cells. The process requires the supply of NADPH.
6. Special functions of NADPH in RBC: NADPH produced in erythrocytes has special functions to perform. It maintains the concentration of reduced glutathione (reaction explained in 3) which is essentially required to preserve the integrity of RBC membrane. NADPH is also necessary to keep the ferrous iron (Fe^{2+}) of hemoglobin in the reduced state so that accumulation of methemoglobin (Fe^{3+}) is prevented.

Glucose 6-phosphate dehydrogenase deficiency

G6PD deficiency is an inherited sex-linked trait. Although the deficiency occurs in all the cells of the affected individuals, it is more severe in RBC.HMP shunt is the only means of providing NADPH in the erythrocytes. Decreased activity of G6PD impairs the synthesis of NADPH in RBC This results in the accumulation of methemoglobin and peroxides in erythrocytes leading to hemolysis.

Clinical manifestations in G6PD deficiency: Most of the patients with G6PD deficiency do not usually exhibit clinical symptoms. Some of them, however, develop **hemolytic anemia** if they are administered oxidant drugs or exposed to a severe infection. The drugs such as primaquine (antimalarial), acetanilide (antipyretic), sulfamethoxazole (antibiotic) or ingestion of fave beans (favism) produce hemolytic jaundice in these patients. Severe infection results in the generation of free radicals (in macrophages) which can enter RBC and cause hemolysis.

G6PD deficiency and resistance to malaria: It is interesting to note that G6PD deficiency is associated with resistance to malaria, caused by *Plasmodium falciparum*. This is explained from the fact that the parasites that cause malaria are dependent on HMP shunt and reduced glutathione for their optimum growth in RBC Therefore, G6PD deficiency-which is seen frequently in Africans - protects them from malaria, a common disease in this region. It is regarded as an adaptability of the people living in malaria-infected regions of the world.

GLYOXYLATE CYCLE

The animals, including man, cannot carry out the net synthesis of carbohydrate from fat. However, the plants and many microorganisms are equipped with the metabolic machinery-namely the glyoxylate cycle-to convert fat into carbohydrates. This pathway is very significant in the germinating seeds where the stored triacylglycerol (fat) is converted to sugars to meet the energy needs.

Location of the cycle: The glyoxylate cycle occurs in glyoxysomes, specialized cellular organelles, where fatty acid oxidation is also operative.

Reactions of the cycle: The glyoxylate cycle (**Fig.**) is regarded as an anabolic variant of citric acid cycle. Acetyl CoA produced. from fatty acid oxidation condenses with oxaloacetate to give citrate which is then converted to isocitrate. At this stage, isocitrate bypasses the citric acid cycle and is cleaved by isocitrate lyase to succinate and glyoxylate. Another molecule of acetyl CoA is now utilized to combine with glyoxylate to form malate. This reaction is catalysed by malate synthase and the malate so formed enters citric acid cycle.

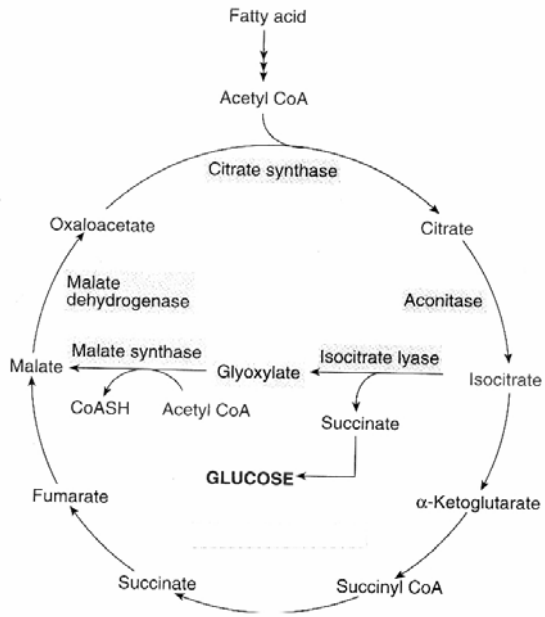


Fig. The glyoxylate cycle

The glyoxylate cycle is a cyclic pathway that results in the conversion of two 2-carbon fragments of acetyl CoA to 4-carbon compound, succinate. The succinate is converted to oxaloacetate and then to glucose involving the reactions of gluconeogenesis.

J5. METABOLISM OF LIPIDS

Lipids are indispensable for cell structure and function. Due to their hydrophobic and non-polar nature, lipids differ from rest of the body compounds and are unique in their action.

Triacylglycerols-the body fuel reserve

Lipids constitute about 15-20% of the body weight in humans. Triacylglycerols (formerly triglycerides) are the most abundant lipids comprising 85-90% of body lipids. Most of the triacylglycerols (TG; also called neutral fat or depot fat) are stored in the adipose tissue and serve as energy reserve of the body. This is in contrast to carbohydrates and proteins which cannot be stored to a significant extent for energy purposes. Fat also acts as an insulating material for maintaining the body temperature of animals.

Why should fat be the fuel reserve of the body?

Triacylglycerols are the most predominant storage form of energy. There are two main reasons for fat being the fuel reserve of the body
1. Triacylglycerols (TG) are highly concentrated form of energy, yielding 9 Cal/g, in contrast to carbohydrates and proteins that produce only 4 Cal/g. This is because fatty acids found in TG are in the reduced form.

2. The triacylglycerols are non-polar and hydrophobic in nature, hence stored in pure form without any association with water (anhydrous form). On the other hand, glycogen and proteins are polar. One gram of glycogen combines with 2 g of water for storage.

For the two reasons stated above, one gram of fat stored in the body yields nearly six times as much energy as one gram of (hydrated) glycogen. In a healthy adult individual (weighing 70 kg), about 10-11 kg of fat is stored (mostly in adipose tissue) which corresponds to a fuel reserve of 100,000 Cals. If this much of energy were to be stored as glycogen (instead of fat), then the weight of the person would increase by at least 55 kg! This explains why fat has been chosen as a fuel reserve during evolution.
The fuel reserve in the form of fat stores will meet the energy requirements for several weeks of food deprivation in man. Hibernating animals provide good example for utilizing fat reserve as fuel.

FATTY ACID OXIDATION

The fatty acids in the body are mostly oxidized by β-oxidation. β-Oxidation may be defined as the **oxidation of fatty acids on the β-carbon atom**. This results in the sequential removal of a two carbon fragment, acetyl CoA.

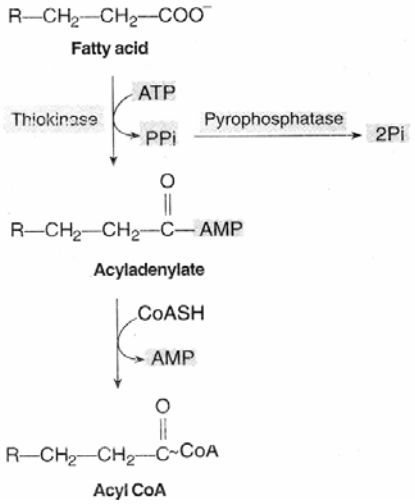
The nature of fatty acid oxidation was identified as early as 1904. Franz Knoop (a German scientist), fed dogs fatty acids tagged (labelled) at their last (ω, omega) carbon with phenyl ring. The dogs excreted in urine a final product phenylacetic acid when fed even carbon fatty acids. For odd carbon fatty acids, benzoic acid was the excretory product. Knoop proposed that fatty acids are degraded by oxidation of the β-carbon. Surprisingly, it took nearly 50 years to confirm Knoop's hypothesis. The enzymes involved in ∼-oxidation and the reaction mechanism were elucidated only by 1950. The prominent biochemists who contributed to our knowledge of fatty acid oxidation include Albert Lehninger, Lewis Leboir, Eugene Kennedy, Feoder Lynen and David Green.

Fatty acid oxidation- stages and tissues

The β-oxidation of fatty acids involves three stages

I. Activation of fatty acids occurring in the cytosol;
II. Transport of fatty acids into mitochondria;
III. β-Oxidation proper in the mitochondrial matrix.

Fatty acids are oxidized by most of the tissues in the body. However, brain, erythrocytes and adrenal medulla cannot utilize fatty acids for energy requirement.



I. Fatty acid activation: Fatty acids are activated to acyl CoA by thiokinases or acyl CoA synthetases. The reaction occurs in two steps and requires ATP, coenzyme A and Mg^{2+}. Fatty acid reacts with ATP to form acyladenylate which then combines with coenzyme A to produce acyl CoA (**Fig.**).

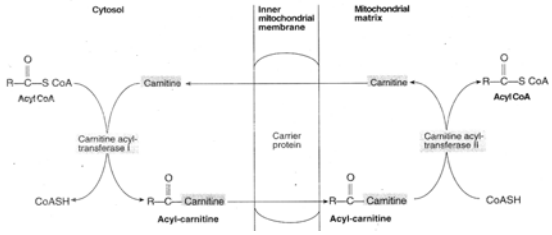
In the activation, two high energy phosphates are utilized, since ATP is converted to pyrophosphate (PPi). The enzyme inorganic pyrophosphatase hydrolyses PPi to phosphate (Pi). The immediate elimination of PPi makes this reaction totally irreversible.

Three different thiokinases, to activate long chain (10-20 carbon), medium chain (4-12 carbon) and short chain <4 carbon) fatty acids have been identified.

II. Transport of acvl CoA into mitochondria

The inner mitochondrial membrane is impermeable to fatty acids. A specialized carnitine carrier system operates to transport activated fatty acids from cytosol to the mitochondria. This occurs in four steps (**Fig.**).

1. Acyl group of acyl CoA is transferred to carnitine (β-hydroxy-γ-trimethyl aminobutyrate), catalysed by the carnitine acyltransferase I (present on the outer surface of inner mitochondrial membrane).
2. The acyl-carnitine is transported across the membrane to mitochondrial matrix by a specific carrier protein.
3. Carnitine acyl transferase II (found on the inner surface of inner mitochondrial membrane) converts acyl-carnitine to acyl CoA.
4. The carnitine released returns to cytosol for reuse.



It should be noted that the coenzyme A used for activation is different from the one that finally combines with fatty acid in the mitochondria to form acyl CoA. Thus, **the cell has two separate pools (cytosolic and mitochondrial) of coenzyme A.**

Inhibitor of carnitine shuttle: Carnitine acyl transferase I is inhibited by malonyl CoA, a key metabolite involved in fatty acid synthesis that occurs in cytosol (details given later). In other words, while the fatty acid synthesis is in progress (reflected by high concentration of malonyl CoA), their oxidation do not occur, since carnitine shuttle is impaired.

III. β-oxldation proper

Each cycle of β-oxidation, liberating a two carbon unit-acetyl CoA, occurs in a sequence of four reactions (Fig.).

1. **Oxidation:** Acyl CoA undergoes dehydrogenation by an FAD-dependent flavoenzyme, acyl CoA dehydrogenase. A double bond is formed between α and β carbons (i.e., 2 and 3 carbons).
2. **Hydration:** Enoyl CoA hydratase brings about the hydration of the double bond to form β-hydroxyacyl CoA.
3. **Oxidation:** β-Hydroxyacyl CoA dehydrogenase catalyses the second oxidation and generates NADH. The product formed is β-ketoacyl CoA.

4. **Cleavage:** The final reaction in β-oxidation is the liberation of a 2 carbon fragment, acetyl CoA from acyl CoA. This occurs by a thiolytic cleavage catalysed by **β-ketoacyl CoA thiolase** (or simply thiolase).

The new acyl CoA, containing two carbons less than the original, reenters the β-oxidation cycle. The process continues till the fatty acid is completely oxidized. The overall reaction for each cycle of β-oxidation

C_n Acyl CoA + FAD + NAD^+ + H_2O + CoASH→ $C_{(n-2)}$ Acyl CoA + Acetyl CoA + $FADH_2$ + NADH + H^+

The scheme of fatty acid oxidation discussed above corresponds to saturated (no double bond) and even carbon fatty acids. This occurs most predominantly in biological system.

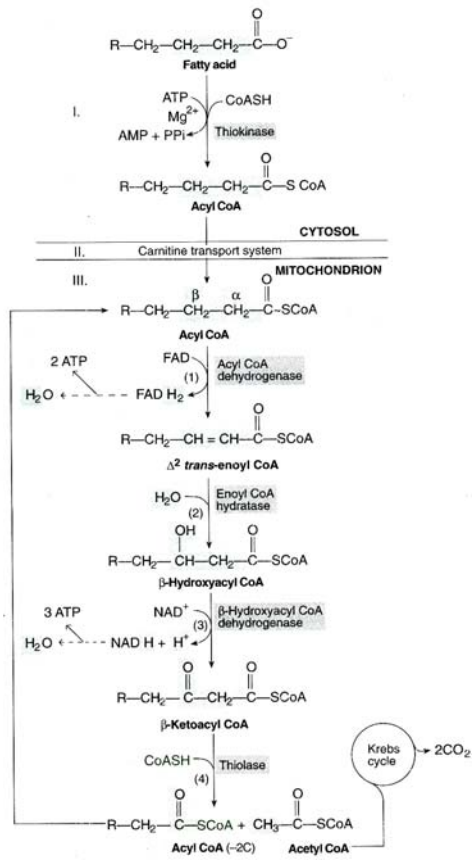
Oxidation of palmitoyl CoA: The summary of β-oxidation of palmitoyl CoA is shown here

Palmitoyl CoA + 7 CoASH + 7 FAD + 7 NAD^++ 7 H_2O→ 8 Acetyl CoA + 7 $FADH_2$ + 7 NADH+ 7H^+

Palmitoyl CoA undergoes 7 cycles of β-oxidation to yield 8 acetyl CoA. Acetyl CoA can enter citric acid cycle and get completely oxidized to CO_2 and H_2O.

Energetics of β-oxidation

The ultimate aim of fatty acid oxidation is to generate energy. The energy obtained from the complete oxidation of palmitic acid (16 carbon) is given in Table.



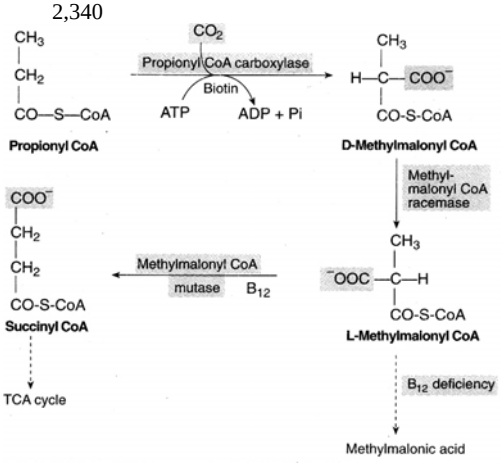
Mechanism	ATP Yield
I. β-Oxidation 7 cycles	
7 $FADH_2$ [Oxidised by electron transport chain, each $FADH_2$ gives 2 ATP]	14
7 NADH [Oxidised by ETC, each NADH liberates 3ATP]	21
II. From 8 Acetyl CoA	
Oxidised by Citric Acid Cycle, each acetyl CoA provides 12 ATP	96
Total energy from one mole of palmitoyl coA	131
Energy utilized for activation (formation of palmitoyl CoA)	- 2
Net yield of Oxidation of one molecule of palmitate	129

The standard free energy of palmitate = 2,340 Cal.
The energy yield by its oxidation-129 ATP (129 x 7.3 Cal) = 940 Cal.
The efficiency of energy conservation by fatty acid oxidation = $\frac{940}{2,340}$ x 100 = 40%.

Oxidation of odd carbon chain fatty acids

The β-oxidation of saturated fatty acids containing odd number of carbon atoms proceeds in the same manner, as described above for even carbon fatty acids. The only difference is that in the last and final β-oxidation cycle, a three-carbon fragment is left behind (in place of 2 carbon unit for saturated fatty acids). This compound is **propionyl CoA** which is converted to succinyl CoA as follows

1. Propionyl CoA is carboxylated in the presence of ATP, CO_2 and vitamin **biotin** to D-methylmalonyl CoA.
2. Methylmalonyl CoA epimerase converts the methylmalonyl CoA to L-form. This reaction (D → L) is essential for the entry of this compound into the metabolic reactions of the body.
3. The next enzyme, methylmalonyl CoA mutase, is dependent on vitamin B_{12} (deoxyadenosyl cobalamin). It catalyses the conversion of methyl malonyl CoA (a branched compound) to succinyl CoA (a straight chain compound), which can enter citric acid cycle.



Oxidation of unsaturated fatty acids:

Due to the presence of double bonds, the unsaturated fatty acids are not reduced to the same extent as saturated fatty acids. Therefore, oxidation of unsaturated fatty acids, in general, provides less energy than that of saturated fatty acids.

Most of the reactions involved in the oxidation of unsaturated fatty acids are the same as found in the β-oxidation of saturated fatty acids. However, the presence of double bonds poses problem for β-oxidation to proceed. This is overcome by two additional enzymes-an isomerase and an epimerase.

β-Oxidation of fatty acids in peroxisomes

Peroxisomes are organelles present in most eukaryotic cells. The **β**-oxidation occurs in a modified form in peroxisomes. Acyl CoA dehydrogenase (a flavoenzyme) leads to the formation of $FADH_2$ as in β-oxidation. The reducing equivalents from $FADH_2$ are not transferred to the electron transport chain, but handed over directly to O_2. This results in the formation of H_2O_2 which is cleaved by catalase.

$$E\text{-}FADH_2 + O_2 \rightarrow E\text{-}FAD + H_2O_2$$

$$H_2O_2 \xrightarrow{\text{Catalase}} H_2O + \tfrac{1}{2} O_2$$

There is no ATP synthesized in peroxisomal **β**-oxidation of fatty acids, since the reducing equivalents do not pass through He. However, heat is liberated.

It is now believed that the peroxisomes carry out the initial oxidation of long chain (C_{20}, C_{22} etc.) fatty acids which is followed by mitochondrial oxidation.

Peroxisomal oxidation is induced by high fat diet and administration of hypolipidemic drugs (e.g. clofibrate).

Zellweger syndrome: This is a rare disorder characterized by the absence of peroxisomes in almost all the tissues. As a result, the long chain fatty acids (C_{26}-C_{38}) are not oxidized. They accumulate in tissues, particularly in brain, liver and kidney. Hence the disorder is also known as **cerebrohepatorenal syndrome**.

α-Oxidation of fatty acids: β-Oxidation is the most predominant pathway for fatty acid degradation. However, the removal of one carbon unit at a time by the oxidation of α-carbon atom of fatty acid is known. α-Oxidation does not involve the binding of fatty acid to coenzyme A and no energy is produced.

ω-Oxidation of fatty acid: This is a minor pathway. It involves hydroxylation followed by oxidation of **ω**-carbon present as a methyl group at the other end (at one end carboxyl group is present) of fatty acid. This reaction requires cytochrome P_{450} NADPH and O_2 besides the enzymes.

BIOSYNTHESIS OF FATTY ACIDS

The dietary carbohydrates and amino acids, when consumed in excess, can be converted to fatty acids and stored as triacylglycerols. De novo (new) synthesis of fatty acids occurs predominantly in liver, kidney, adipose tissue and lactating mammary glands. The enzyme machinery for fatty acid production is located in the cytosomal fraction of the cell. Acetyl CoA is the source of carbon atoms while NADPH provides the reducing equivalents and ATP supplies energy for fatty acid formation. The fatty acid synthesis may be learnt in 3 stages

I. Production of acetyl CoA and NADPH
II. Conversion of acetyl CoA to malonyl CoA
III. Reactions of fatty acid synthase complex.

I. Production of acetyl CoA and NADPH: Acetyl CoA and NADPH are the prerequisites for fatty acid synthesis. Acetyl CoA is produced in the mitochondria by the oxidation of pyruvate and fatty acids, degradation of carbon skeleton of amino acids, and from ketone bodies. Mitochondria, however, are not permeable to acetyl CoA. An alternate or a bypass arrangement is made for the transfer of acetyl CoA to cytosol. Acetyl CoA condenses with oxaloacetate in mitochondria to form citrate. Citrate is freely transported to cytosol where it is cleaved by **citrate lyase** to liberate acetyl CoA and oxaloacetate. Oxaloacetate in the cytosol is converted to malate. Malic enzyme converts malate to pyruvate. NADPH and CO_2 are generated in this reaction. Both of them are utilized for fatty acid synthesis.

Advantages of coupled transport of acetyl CoA and NADPH : The transport of acetyl CoA from mitochondria to cytosol is coupled with the cytosomal production of NADPH and CO_2 which is highly advantageous to the cell for optimum synthesis of fatty acids. For the synthesis of palmitate, 8 acetyl CoA are transported from the mitochondria to cytosol which is linked with the synthesis of 8 NADPH in the cytosol. As such, 14 NADPH are needed for the production of one molecule of palmitate. The remaining 6 NADPH, therefore, must be obtained from HMP shunt. This sort of calculation may be more hypothetical than real, since many metabolic reactions simultaneously operate in the cell.

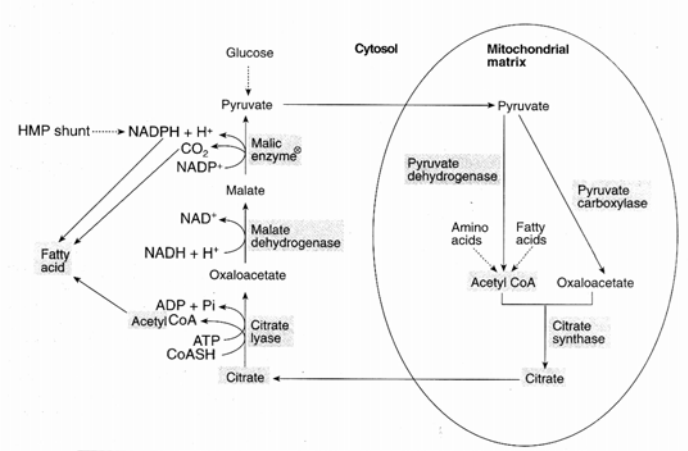
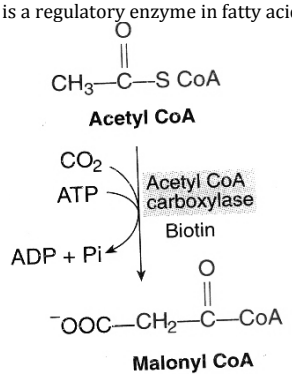


Fig. : Transfer of acetyl CoA from mitochondria to cytosol (HMP shunt—Hexose monophosphate shunt; ⊗—also known as malate dehydrogenase).

II. Formation of malonyl CoA: Acetyl CoA is carboxylated to malonyl CoA by the enzyme acetyl CoA carboxylase. This is an ATP-dependent reaction and requires biotin for CO_2 fixation. The mechanism of action of acetyl CoA carboxylase is similar to that of pyruvate carboxylase. Acetyl CoA carboxylase is a regulatory enzyme in fatty acid synthesis.

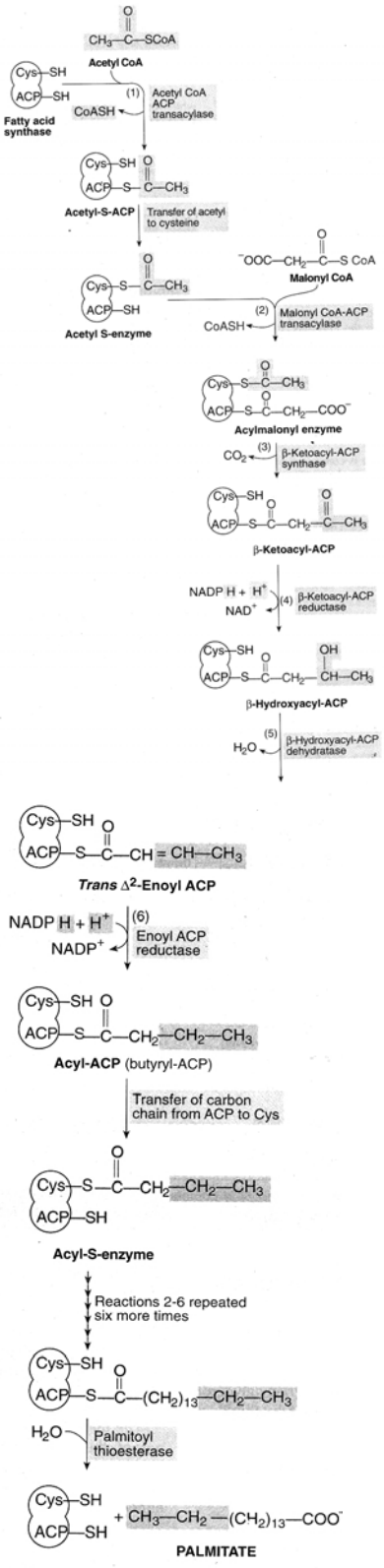


III. Reactions of fatty acid synthase complex

The remaining reactions of fatty acid synthesis are catalysed by a multifunctional enzyme known as fatty acid synthase (FAS) complex. In eukaryotic cells, including man, the fatty acid synthase exists as a dimer with two identical units. Each monomer possesses the activities of seven different enzymes and an acyl carrier protein (ACP) bound to 4'-phosphopantetheine. Fatty acid synthase functions as a single unit catalysing all the seven reactions. Dissociation of the synthase complex results in loss of the enzyme activities. In the lower organisms (prokaryotes), the fatty acid synthesis is carried out by a multienzyme complex in association with a separate acyl carrier protein. This is in contrast to eukaryotes where ACP is a part of fatty acid synthase.

The sequence of reactions for the extra-mitochondrial synthesis of fatty acids (palmitate) is depicted in **Fig** and described below

1. The two carbon fragment of acetyl CoA is transferred to ACP of fatty acid synthase, catalysed by the enzyme, **acetyl CoA-ACP transacylase**. The acetyl unit is then transferred from ACP to cysteine residue of the enzyme. Thus ACP site falls vacant.
2. The enzyme **malonyl CoA-ACP transacylase** transfers malonate from malonyl CoA to bind to ACP.
3. The acetyl unit attached to cysteine is transferred to malonyl group (bound to ACP). The malonyl moiety loses CO_2 which was added by **acetyl CoA carboxylase**. Thus, CO_2 is never incorporated into fatty acid carbon chain. The decarboxylation is accompanied by loss of free energy which allows the reaction to proceed forward. This reaction is catalyzed by **β-ketoacyl ACP synthase**.
4. β-Ketoacyl ACP reductase reduces ketoacyl group to hydroxyacyl group. The reducing equivalents are supplied by NADPH.
5. β-Hydroxyacyl ACP undergoes dehydration. A molecule of water is eliminated and a double bond is introduced between a. and 13 carbons.
6. A second NADPH-dependent reduction, catalysed by **enoyl-ACP reductase** occurs to produce acyl-ACP. The four-carbon unit attached to ACP is butyryl group.The carbon chain attached to ACP is transferred to cysteine residue and the reactions 2-6 are repeated 6 more times. Each time, the fatty acid chain is lengthened by a two-carbon unit (obtained from malonyl CoA). At the end of 7 cycles, the fatty acid synthesis is complete and a 16-carbon fully saturated fatty acid-namely palmitate-bound to ACP is produced.
7. The enzyme palmitoyl thioesterase separates palmitate from fatty acid synthase. This completes the synthesis of palmitate.



Summary of palmitate synthesis

Of the 16 carbons present in palmitate, only two come from acetyl CoA directly. The remaining 14 are from malonyl CoA which, in turn, is produced by acetyl CoA. The overall reaction of palmitate synthesis is summarized

8 Acetyl CoA + 7 ATP + 14 NADPH + 14 H^+ → Palmitate + 8 CoA + 7 ADP + 7 Pi + 6H_2O

Fatty acid synthase complex

The diagrammatic representation of the model for fatty acid synthase (FAS) multienzyme complex is depicted in Fig. This model is tentative and is largely based on the work of Wakil.

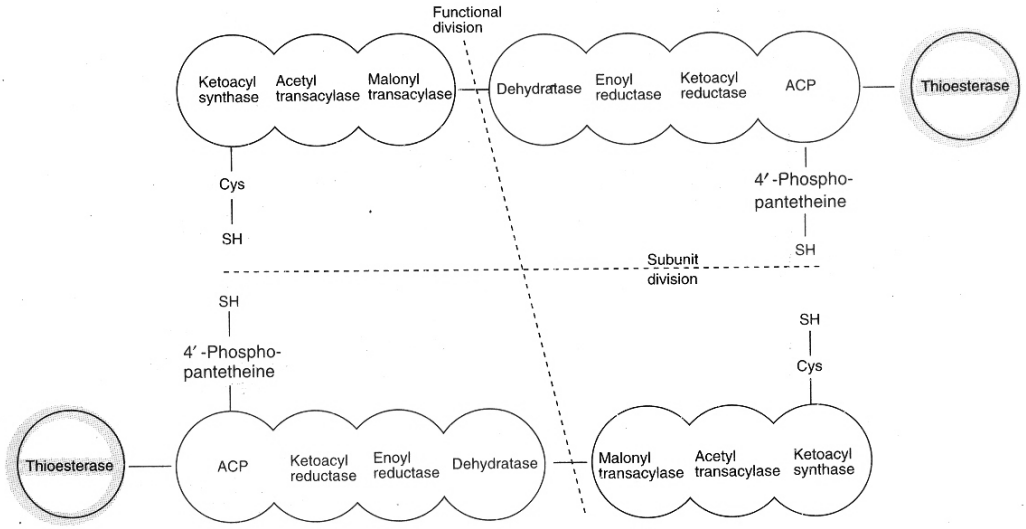
Fatty acid synthase is a **dimer** composed of two identical subunits (monomers), each with a molecular weight of 240,000. Each subunit contains the activities of 7 enzymes of FAS and an ACP with 4'-phosphopantetheine -SH group. The two subunits lie in anti parallel (head-to-tail) orientation. The -SH group of phosphopantetheine of one subunit is in close proximity to the -SH of cysteine residue (of the enzyme ketoacyl synthase) of the other subunit.

Each monomer of FAS contains all the enzyme activities of fatty acid synthesis. But only the dimer is functionally active. This is because the functional unit consists of half of each subunit interacting with the complementary half of the other. Thus, the FAS structure has both functional division and subunit division . The two functional subunits of FAS independently operate and synthesize two fatty acids simultaneously.

Functional significance of FAS complex

The organization of different enzymes of a metabolic pathway into a single multienzyme functional unit has distinct advantages for cellular function

1. The FAS complex offers great efficiency that is free from interference of other cellular reactions for the synthesis of fatty acids.
2. Since the entire process of the metabolic pathway is confined to the complex, there are no permeability barriers for the various intermediates.
3. The multienzyme polypeptide complex is coded by a single gene. Thus, there is a good coordination in the synthesis of all enzymes of the FAS complex.



Regulation of fatty acid synthesis

Fatty acid production is controlled by enzymes, metabolites, end products, hormones and dietary manipulations. Some of the important regulatory mechanisms are discussed hereunder.

Acteyl CoA carboxylase: This enzyme controls a committed step in fatty acid synthesis. Acetyl CoA carboxylase exists as an inactive protomer (monomer) or an active polymer. Citrate promotes polymer formation, hence increases fatty acid synthesis. On the other hand, palmitoyl CoA and malonyl CoA cause depolymerization of the enzyme and, therefore, inhibit fatty acid synthesis.

Hormonal influence: Hormones regulate acetyl CoA carboxylase by a separate mechanismphosphorylation (inactive form) and dephosphorylation (active form) of the enzyme. Glucagon, epinephrine and norepinephrine inactivate the enzyme by cAMP-dependent phosphorylation. Insulin, on the other hand, dephosphorylates and activates the enzyme. Thus, **insulin promotes fatty acid synthesis while glucagon inhibits**.

Insulin stimulates tissue uptake of glucose, and conversion of pyruvate to acetyl CoA. This also facilitates fatty acid formation.

Dietary regulation: Consumption of high carbohydrate or fat-free diet increases the synthesis of acetyl CoA carboxylase and fatty acid synthase, which promote fatty acid formation. On the other hand, fasting or high fat diet decrease fatty acid production by reducing the synthesis of these two enzymes.

Availability of NADPH: The reducing equivalents for fatty acid synthesis are provided by NADPH which come either from citrate (acetyl CoA) transport or hexose monophosphate shunt. About 50-60% of required NADPH is obtained from HMP shunt, which influences fatty acid synthesis.

Desaturatlon of fatty acid chains: A microsomal enzyme system called **fatty acyl CoA desaturase** is responsible for the formation of unsaturated fatty acids. This reaction also involves flavin-dependent cytochrome bs reductase, NADH and molecular 02. The monounsaturated fatty acids-namely oleic acid and palmitoleic acidare, respectively, synthesized from stearate and palmitate.

Mammals lack the ability to introduce double bonds in fatty acids beyond carbons 9 and 10. Hence, linoleic acid (18: 2; 9, 12) and linolenic acid (18: 3; 9, 12, 15) are essential for man in the diet. However, arachidonic acid (20: 4; 5, 8, 11, 14) can be synthesized from linoleic acid by desaturation and chain elongation. Arachidonic acid is the precursor for eicosanoids (prostaglandins and thromboxanes), a group of compounds with diversified functions.

SYNTHESIS OF LONG CHAIN FATTY ACIDS FROM PALMITATE

Palmitate is the end product of the reactions of fatty acid synthase system that occurs in cytosol. Further, **chain elongation** can take place either in mitochondria or in **endoplasmic reticulum** (microsomes), by separate mechanisms. The microsomal chain elongation is more predominant and involves successive additions of malonyl CoA with the participation of NADPH. These reactions are similar to that catalysed by fatty acid synthase.

A specific group of enzymes, namely elongases, bring about fatty acid chain elongation.

The mitochondrial chain elongation is almost a reversal of **β**-oxidation of fatty acids. Acetyl CoA are successively added to fatty acid to lengthen the chain. The reducing equivalents are derived from NADPH.

Comparison between fatty acid synthesis and oxidation

The synthesis of fatty acids and their oxidation are two distinct and independent pathways. A comparison between these two metabolic pathways in given in Table.

		fatty acid synthesis	β-Oxidation
1.	Major tissues	Liver, adipose tissue	Muscle, liver
2.	Subcellular site	Cytosol	Mitochondria
3.	Precursor/substrate	Acetyl CoA	Acyl CoA
4.	End product	Palmitate	Acetyl CoA
5.	Intermediates are bound to	Acyl carrier protein	Coenzyme A
6.	Coenzyme requirement	NADPH (supplying reducing equivalents)	FAD and NAD^+ (get reduced)
7.	Carbon units added/degraded	Malonyl CoA	Acetyl CoA
8.	Transport system	Citrate (mitochondria cytosol)	Carnitine cytosol mitochondria)
9.	Inhibitor	Long chain acyl CoA (inhibits acetyl CoA carboxylase)	Malonyl CoA (inhibits carnitine acyl transferase I)
10.	The pathway increased	After rich carbohydrate diet	In starvation
11.	Hormonal status that promotes	High ratio of insulin/glucagon	Low ratio of insulin/glucagon
12.	Status of enzyme(s)	Multifunctional enzyme complex	Individual enzymes

SYNTHESIS OF TRIACYLGLYCEROLS

Triacylglycerol (TG) synthesis mostly occurs in liver and adipose tissue, and to a lesser extent in other tissues. Fatty acids and glycerol must be activated prior to the synthesis of triacylglycerols. Conversion of fatty acids to acyl CoA by thiokinase is already described .

Synthesis of glycerol 3-phosphate

Two mechanisms are involved for the synthesis of glycerol 3-phosphate

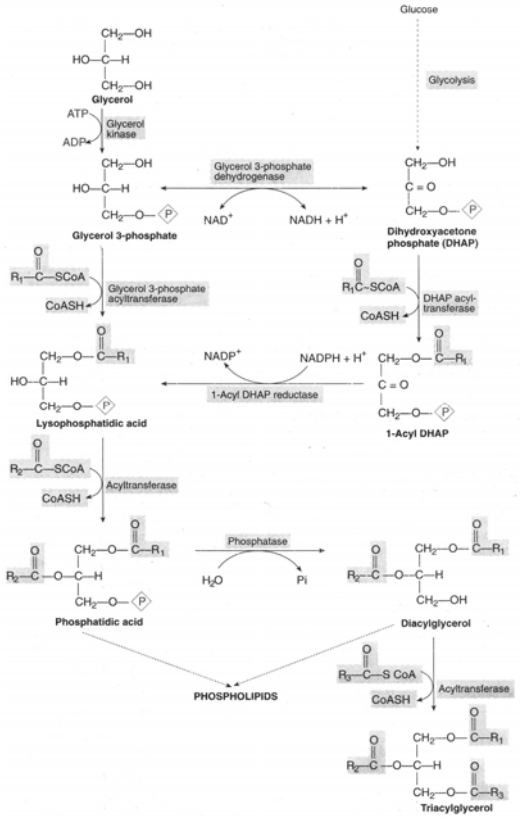
1. In the liver, glycerol is activated by glycerol kinase. This enzyme is absent in adipose tissue.
2. In both liver and adipose tissue, glucose serves as a precursor for glycerol 3-phosphate. Dihydroxyacetone phosphate (DHAP) produced in glycolysis is reduced by glycerol 3-phosphate dehydrogenase to glycerol 3-phosphate.

Addition of acyl groups to form TG

Glycerol 3-phosphate acyltransferase catalyses the transfer of an acyl group to produce lysophosphatidic acid. DHAP can also accept acyl group, ultimately resulting in the formation of lysophosphatidic acid. Another acyl group is added to lysophosphatidic acid to form phosphatidic acid (1, 2-diacylglycerol phosphate). The enzyme phosphatase cleaves off phosphate of phosphatidic acid to produce diacylglycerol. Incorporation of another acyl group finally results in synthesis of triacylglycerol.

The three fatty acids found in triacylglycerol are not of the same type. A saturated fatty acid is usually present on carbon 1, an unsaturated fatty acid is found on carbon 2 and carbon 3 may have either.

The intermediates of TG synthesis phosphatidic acid and diacylglycerol are also utilized for phospholipid synthesis.



METABOLISM OF PHOSPHOLIPIDS

Phospholipids are a specialized group of lipids performing a variety of functions. These include the membrane structure and functions, involvement in blood clotting and supply of arachidonic acid for the synthesis of. prostaglandins.

Synthesis of phospholipids

Phospholipids are synthesized from phosphatidic acid and 1, 2-diacylglycerol, intermediates in the production of triacylglycerols (**Fig.**). Phospholipid synthesis occurs in the smooth endoplasmic reticulum.

1. Formation of lecithin and cephalin: Choline and ethanolamine first get phosphorylated and then combine with CTP to form, respectively, COPcholine and COP-ethanolamine. Phosphatidylcholine (lecithin) is synthesized when COP-choline combines with 1, 2-diacylglycerol. Phosphatidyl ethanolamine (cephalin) is produced when COP-ethanolamine reacts with 1, 2-diacylglycerol. Phosphatidyl ethanolamine can be converted to phosphatidyl choline on methylation.

2. Synthesis of phosphatidylserine: Phosphatidyl ethanolamine can exchange its ethanolamine group with free serine to produce phosphatidyl serine. The latter, on decarboxylation, gives the former.

3. Formation of phosphatidylinositol: COPdiacylglycerol produced from phosphatidic acid combines with inositol to form phosphatidyl inositol (PI). This phospholipid contains arachidonic acid on carbon 2 of glycerol which serves as a substrate for prostaglandin synthesis. Further, Pi is important for signal transmission across membranes.

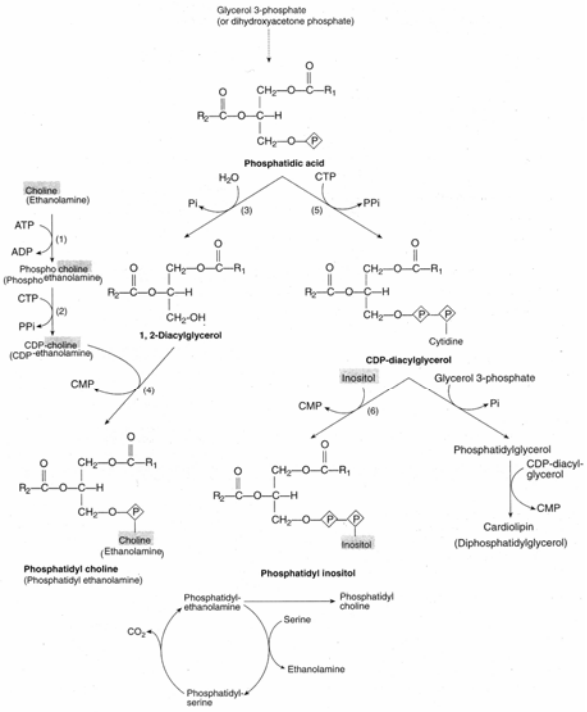


Fig. : Biosynthesis of phospholipids [The enzymes are numbered—(1) Choline kinase, (2) Phosphocholine cytidyltransferase, (3) Phosphatidate phosphohydrolase, (4) Phosphocholine diacylglycerol transferase, (5) CTP—Phosphatidate cytidyltransferase, (6) CDP—Diacylglycerol inositol transferase].

4. Synthesis of phosphatidyl glycerol and cardiolipin: COP-diacylglycerol combines with glycerol 3-phosphate to form phosphatidyl glycerol which, in turn, condenses with another molecule of COP-diacylglycerol to produce cardiolipin (diphosphatidyl glycerol). Cardiolipin is the only phospholipid possessing antigenic properties.

5. Formation of plasmalogens: These are phospholipids with fatty acid at carbon 1 bound by an ether linkage instead of ester linkage. An important plasmalogen-l-alkenyl 2-acetyl glycerol 3-phosphocholine, causes blood platelet aggregation and is referred to as platelet-activating factor (PAF).

6. Synthesis of sphingomyelins: These are phospholipids containing a complex amino alcohol, sphingosine, instead of glycerol. Palmitoyl CoA and serine combine and undergo a sequence of reactions to produce sphingosine which is then acylated to produce ceramide. Sphingomyelin is synthesized when ceramide combines with CDPcholine .

Degradation of phospholipids

Phospholipids are .degraded by phospholipases which cleave the phosphodiester bonds. These enzymes are found in mammalian tissues, pancreatic juice, snake venom and in some toxins. Certain pathogenic bacteria produce phospholipases which help in the spread of infection by dissolving cell membranes. The important phospholipases and their respective sites of hydrolysis are given in Fig.

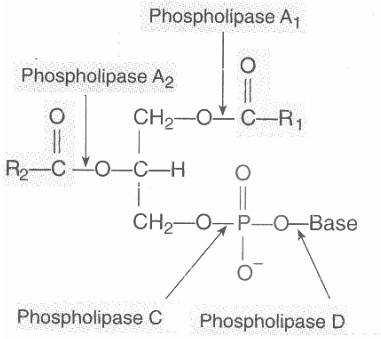


***Fig.:** Degradation of phospholipids by phospholipases.*

- **Phospholipase A_1** specifically cleaves the fatty acid at C1 position of phospholipids resulting in lysophospholipid. The latter can be further acted by lysophospholipase, **phospholipase B** to remove the second acyl group at C_2 position.
- **Phospholipase A_2** hydrolyses the fatty acid at C_2 position of phospholipids. **Snake venom and bee venom** are rich sources of phospholipase A_2. This enzyme is found in many tissues and pancreatic juice. Phospholipase A_2 acts on phosphatidyl inositol to liberate arachidonic acid, the substrate for the synthesis of prostaglandins.
- **Phospholipase C** specifically cleaves the bond between phosphate and glycerol of phospholipids. This enzyme is present in lysosomes of hepatocytes. The toxins isolated from clostridia and other bacilli contain phospholipase C.
- **Phospholipase D** hydrolyses and removes the nitrogenous base from phospholipids. This enzyme is mostly found in plant sources (cabbage, cotton seed etc.).

The degraded products of phospholipids enter the metabolic pool and are utilized for various purposes.

METABOLISM OF CHOLESTEROL

Cholesterol is found exclusively in animals; hence it is often called as animal sterol. The total body content of cholesterol in an adult man weighing 70 kg is about 140 g i.e., around 2 g/kg body weight. Cholesterol is **amphipathic** in nature, since it possesses both hydrophilic and hydrophobic regions in the structure.

Functions of cholesterol:

Cholesterol is essential to life, as it performs a number of important functions

1. It is a structural component of cell membrane.
2. Cholesterol is the precursor for the synthesis of all other steroids in the body. These include steroid hormones, vitamin D and bile acids.
3. It is an essential ingredient in the structure of lipoproteins in which form the lipids in the body are transported.
4. Fatty acids are transported to liver as cholesteryl esters for oxidation.

CHOLESTEROL BIOSYNTHESIS

About 1 g of cholesterol is synthesized per day in adults. Almost all the tissues of the body participate in cholesterol biosynthesis. The largest contribution is made by **liver** (50%), intestine (15%), skin, adrenal cortex, reproductive tissue etc.

The enzymes involved in cholesterol synthesis are found in the **cytosol and microsomal fractions** of the cell. Acetate of acetyl CoA provides all the carbon atoms in cholesterol. The reducing equivalents are supplied by NADPH while ATP provides energy. By administering acetate with ^{14}C isotope label either on the methyl ($-CH_3$) group or carboxyl ($-COO$) group, the origin of carbon atoms in the entire molecule of cholesterol has been established (Konrad Bloch, 1940). The sources of carbon atoms and the key intermediates of cholesterol formation are depicted in Fig.

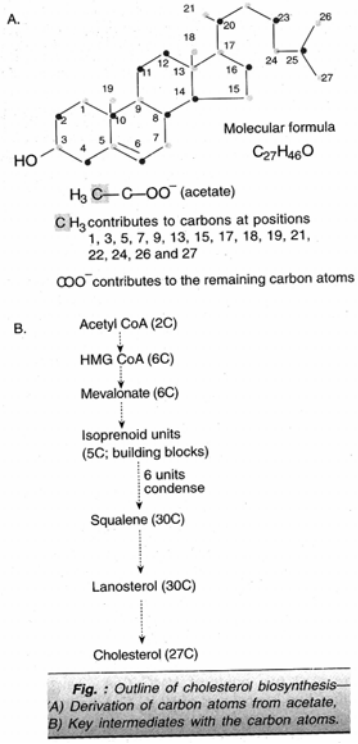


Fig. : Outline of cholesterol biosynthesis— A) Derivation of carbon atoms from acetate, B) Key intermediates with the carbon atoms.

The synthesis of cholesterol may be learnt in 5 stages

1. Synthesis of HMG CoA
2. Formation of mevalonate (6C)
3. Production of isoprenoid units (5C)
4. Synthesis of squalene (30C)
5. Conversion of squalene to cholesterol (27C)

1. Synthesis of β-hydroxy β-methylglutaryl CoA (HMG CoA):

Two moles of acetyl CoA condense to form acetoacetyl CoA. Another molecule of acetyl CoA is then added to produce HMG CoA. These reactions are similar to that of ketone body synthesis. However, the two pathways are distinct, since ketone bodies are produced in mitochondria while cholesterol synthesis occurs in cytosol. Thus, there exist two pools of HMG CoA in the cell. Further, two isoenzymes of HMG CoA synthase are known. The cytosomal enzyme is involved in cholesterol synthesis whereas the mitochondrial HMG CoA' synthase participates in ketone body formation.

2. Formation of mevalonate:

HMG CoA reductase is the rate limiting enzyme in cholesterol biosynthesis. This enzyme is present in endoplasmic reticulum and catalyses the reduction of HMG CoA to mevalonate. The reducing equivalents are supplied by NADPH.

3. Production of isoprenoid units:

In a three step reaction catalysed by kinases, mevalonate is converted to 3-phospho 5-pyrophosphomevalonate which on decarboxylation forms isopentenyl pyrophosphate (lPP). The latter isomerizes to dimethylallyl pyrophosphate (DPP). Both IPP and DPP are 5-carbon isoprenoid units.

4. Synthesis of squalene: IPP and DPP condense to produce a 10-carbon geranyl pyrophosphate (GPP). Another molecule of IPP condenses with GPP to form a 15-carbon farnesyl pyrophosphate (FPP). Two units of farnesyl pyrophosphate unite and get reduced to produce a 3D-carbon squalene.

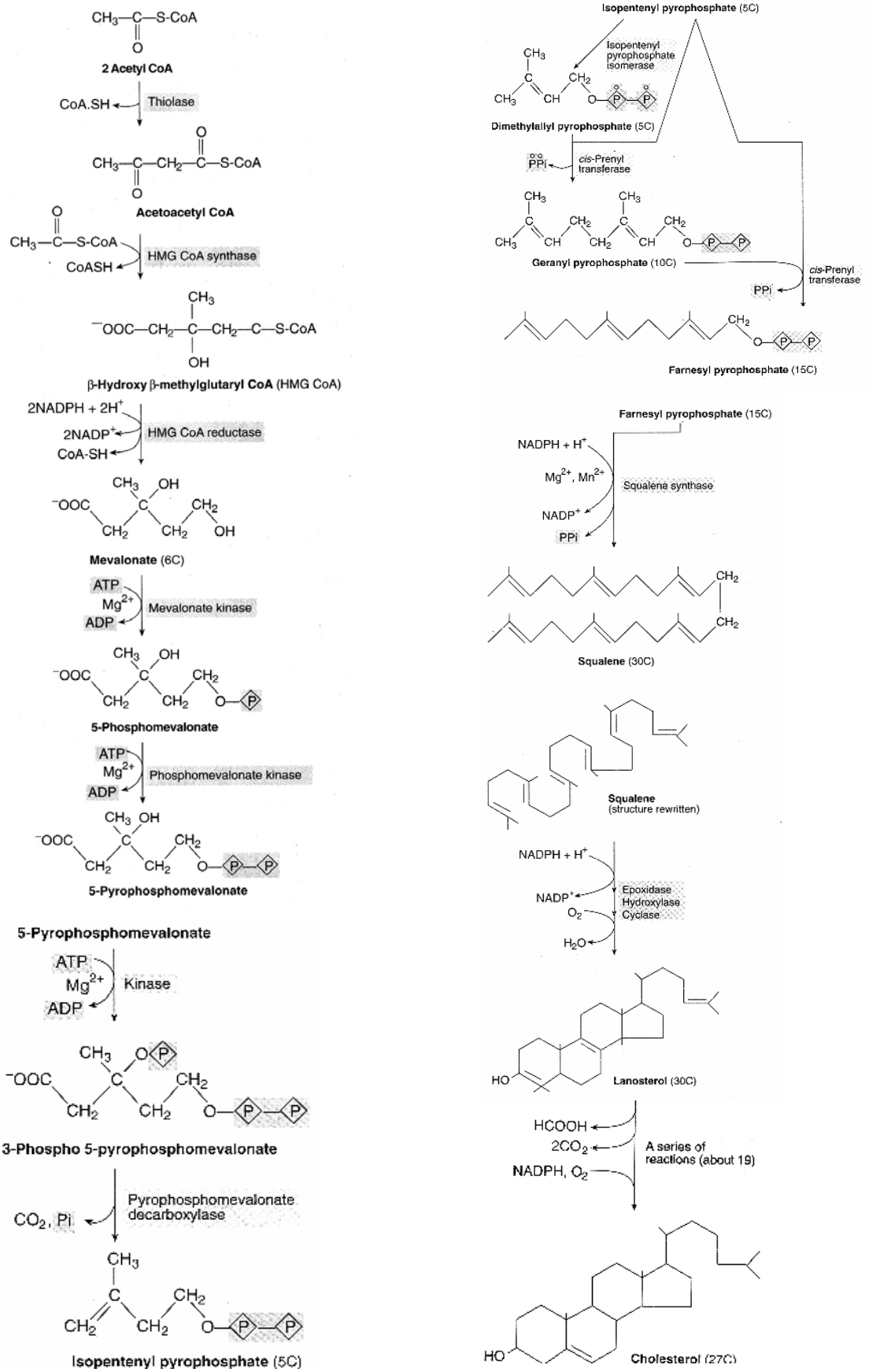
5. Conversion of squalene to cholesterol:

Squalene undergoes hydroxylation and cyclization utilizing 02 and NADPH and gets converted to lanosterol. The formation of cholesterol from lanosterol is a multistep process with a series of about 19 enzymatic reactions. The following are the most important reactions

a) Reducing the carbon atoms from 30 to 27.
b) Removal of two methyl groups from C4 and one methyl group from C_{14}
c) Shift of double bond from C_8 to C_5
d) Reduction in the double bond present between C_{24} and C_{25}

The enzymes involved for the conversion of lanosterol to cholesterol are associated with endoplasmic reticulum. 14-Desmethyl lanosterol, zymosterol, cholestadienol and desmosterol are among the intermediates in the cholesterol biosynthesis. **The penultimate product is 7-dehydrocholesterol which, on reduction, finally yields cholesterol.**

Cholesterol biosynthesis is now believed to be a part of a major metabolic pathway concerned with the synthesis of several other isoprenoid compounds. These include ubiquinone (coenzyme Q of electron transport chain)' and dolichol (found in glycoprotein). Both of them are derived from farnesyl pyrophosphate.


2 Acetyl CoA
Thiolase
Acetoacetyl CoA
HMG CoA synthase
β-Hydroxy β-methylglutaryl CoA (HMG CoA)
HMG CoA reductase
Mevalonate (6C)
Mevalonate kinase
5-Phosphomevalonate
Phosphomevalonate kinase
5-Pyrophosphomevalonate
5-Pyrophosphomevalonate
Kinase
3-Phospho 5-pyrophosphomevalonate
Pyrophosphomevalonate decarboxylase
Isopentenyl pyrophosphate (5C)
Isopentenyl pyrophosphate (5C)
Isopentenyl pyrophosphate isomerase
Dimethylallyl pyrophosphate (5C)
cis-Prenyl transferase
Geranyl pyrophosphate (10C)
cis-Prenyl transferase
Farnesyl pyrophosphate (15C)
Farnesyl pyrophosphate (15C)
Squalene synthase
Squalene (30C)
Squalene
(structure rewritten)
Epoxidase
Hydroxylase
Cyclase
Lanosterol (30C)
A series of reactions (about 19)
Cholesterol (27C)

Regulation of cholesterol synthesis

Cholesterol biosynthesis is controlled by the rate limiting enzyme HMG CoA reductase, at the beginning of the pathway. HMG CoA reductase is found in association with endoplasmic reticulum and is subjected to different metabolic controls.

1. Feedback control: The end product cholesterol controls its own synthesis by a feedback mechanism. Increase in the cellular concentration of cholesterol reduces the synthesis of the enzyme HMG CoA reductase. This is achieved by decreasing the transcription of the gene responsible for the production of HMG CoA reductase. Feedback regulation has been investigated with regard to LDL-cholesterol taken up by the cells and
the same mechanism is believed to operate whenever cellular cholesterol level is elevated.

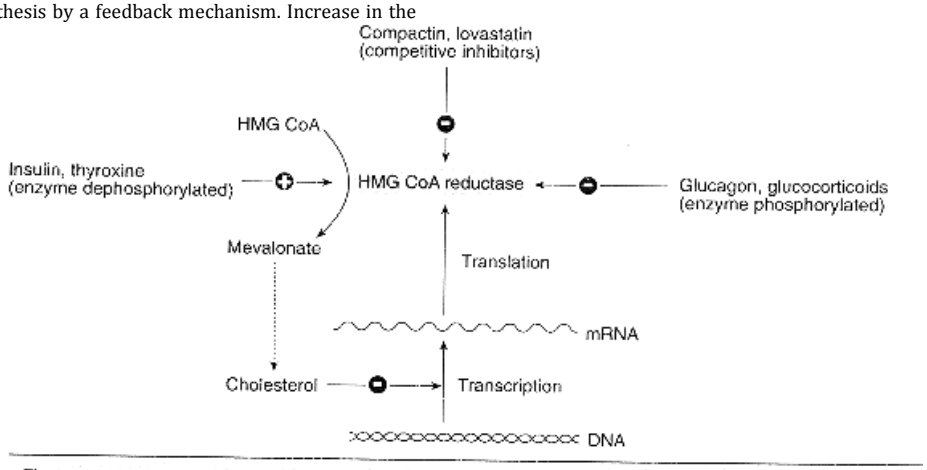


Fig. : Regulation of cholesterol biosynthesis by HMG CoA reductase (⊕—Promoting effect; ⊖—Inhibitory effect).

2. Hormonal regulation: The enzyme HMG CoA reductase exists in two interconvertible forms. The dephosphorylated form of HMG CoA is more active while the phosphorylated form is less active. The hormones exert their influence through cAMP by a series of reactions which are comparable with the control of the enzyme glycogen synthase. The net effect is that glucagon and glucocorticoids favour the formation of inactive HMG CoA reductase (phosphorylated form) and, thus, decrease cholesterol synthesis. On the other hand, insulin and thyroxine increase cholesterol production by enhancing the formation of active HMG CoA reductase (dephosphorylated form).

3. Inhibition by drugs: The drugs compactin and lovastatin (mevinolin) are fungal products. They are used to decrease the serum cholesterol level in patients with hypercholesterolemia. Compactin and lovastatin are competitive inhibitors of the enzyme HMG CoA reductase and, therefore, reduce cholesterol synthesis. About 50 to 60% decrease in serum cholesterol level has been reported by a combined use of these two drugs.

4. HMG CoA reductase activity is inhibited by **bile acids**. Fasting also reduces the activity of this enzyme.

DEGRADATION OF CHOLESTEROL

The steroid nucleus (ring structure) of the cholesterol cannot be degraded to CO_2 and H_2O. Cholesterol (50%) is converted to bile acids, excreted in feces, serves as a precursor for the synthesis of steroid hormones, vitamin D, coprostanol and cholestanol. The latter two are the fecal sterols, besides cholesterol.

I. Synthesis of bile acids: The bile acids possess 24 carbon atoms, 2 or 3 hydroxyl groups in the steroid nucleus and a side chain ending in carboxyl group. The bile acids are amphipathic in nature since they possess both polar and non-polar groups. They serve as emulsifying agents in the intestine and actively participate in the digestion and absorption of lipids.

The synthesis of primary bile acids takes place in the liver and involves a series of reactions (Fig.). The step catalysed by **7 α-hydroxylase** is inhibited by bile acids and this is the rate limiting reaction. Cholic acid and chenodeoxycholic acid are the primary bile acids and the former is found in the largest amount in bile. On conjugation with glycine or taurine, conjugated bile acids (glycocholic acid, taurocholic acid etc.) are formed which are more efficient in their function as surfactants. In the bile, the conjugated bile acids exist as sodium and potassium salts which are known as bile salts.

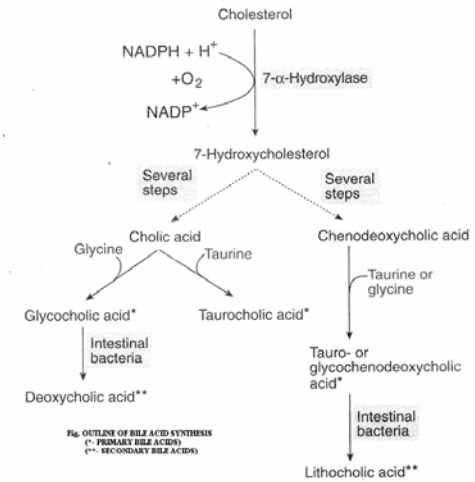


Fig. OUTLINE OF BILE ACID SYNTHESIS (*- PRIMARY BILE ACIDS) (**- SECONDARY BILE ACIDS)

In the intestine, a portion of primary bile acids undergo deconjugation and dehydroxylation to form secondary bile acids (deoxycholic acid and lithocholic acid). These reactions are catalysed by bacterial enzymes in the synthesized in the liver accumulate in gall bladder. From there they are secreted into the small intestine where they serve as emulsifying agents for the digestion and absorption of fats and fat soluble vitamins. A large portion of the bile salts (primary and secondary) are intestine.

Enterohepatic circulation: The conjugated bile salts reabsorbed and returned to the liver through portal vein. Thus the bile salts are recycled and reused several times in a

day. This is known as enterohepatic circulation. About 15-30 g of bile salts are secreted into the intestine each day and reabsorbed. However, a small portion of about 0.5 g/day is lost in the feces. An equal amount (0.5 glday) is synthesized in liver to replace the lost bile salts. The fecal excretion of bile salts is the only route for the removal of cholesterol from the body.

Cholelithiasis: Bile salts and phospholipids are responsible for keeping the cholesterol in bile in a soluble state. Due to their deficiency (particularly bile salts), cholesterol crystals precipitate in the gall bladder often resulting in cholelithiasis-cho/esterol gall stone disease. Cholelithiasis may be due to defective absorption of bile salts from the intestine, impairment in liver function, obstruction of biliary tract etc.

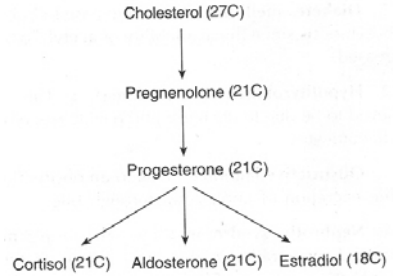
The patients of cholelithiasis respond to the administration of bile acid chenodeoxy cholic acid, commonly known as chenodiol. It is believed that a slow but gradual dissolution of gall stones occursdue to chenodiol. For severe cases of cholelithiasis, surgical removal of gall bladder is the only remedy.

II. Synthesis of steroid hormones from cholesterol

Cholesterol is the precursor for the synthesis of all the five classes of steroid hormones

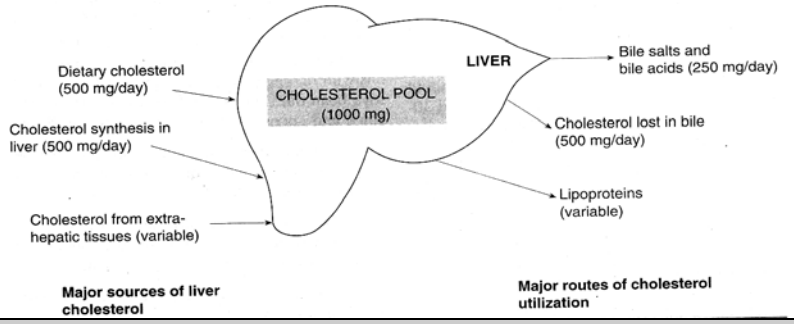
a) Glucocorticoids (e.g. cortisol)
b) Mineralocorticoids (e.g. aldosterone)
c) Progestins (e.g. progesterone)
d) Androgens (e.g. testosterone)
e) Estrogens (e.g. estradiol).

A brief outline of steroid hormonal synthesis is given in Fig.



III. Synthesis of Vitamin D: 7-hydroxychloesterol, an intermediate in the synthesis of cholesterol, is converted to cholecalciferol (Vitamin D_3) by UV rays in skin.

A brief summary of prominent sources and major pathways for utilization of cholesterol with the liver as central metabolic organ is depicited in figure.



J6. METABOLISM OF AMINO ACIDS

The amino acids obtained from dietary source or body protein turnover are utilized for protein biosynthesis and the production of a wide range of nitrogen-containing compounds (Creatine, amines, porphyrin etc.)

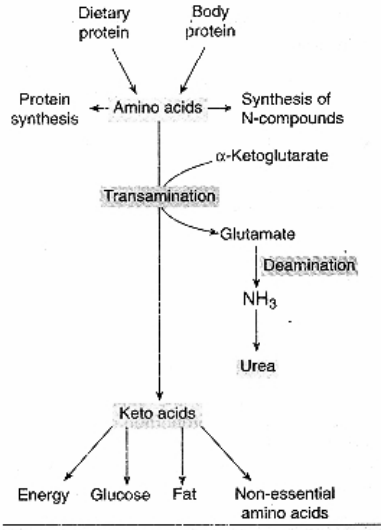
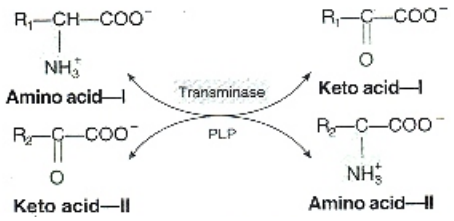


Fig. : An overview of amino acid metabolism.

The amino acids undergo certain common reactions like transamination followed by deamination for the liberation of ammonia. The amino group of the amino acids is utilized for the formation of urea which is an excretory end product of protein metabolism. The carbon skeleton of the amino acids is first converted to keto acids (by transamination) which meet one or more of the following fates.
1. Utilized to generate energy.
2. Used for the synthesis of glucose.
3. Diverted for the formation of fat or ketone bodies.
4. Involved in the production of non-essential amino acids.

The details of general and specific metabolic reactions of amino acids are described in the following pages.

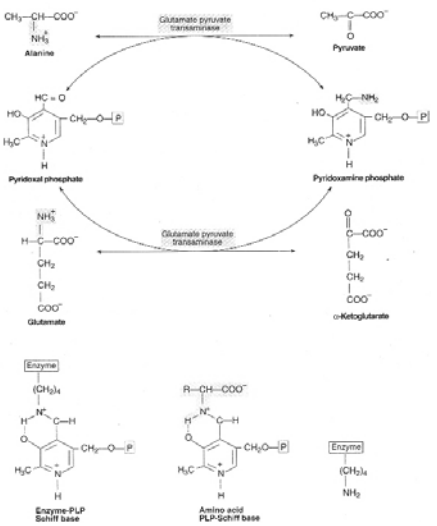
TRANSAMINATION



The transfer of an amino (-NH_2) group from an amino acid to a keto acid is known as transamination. This process involves the interconversion of a pair of amino acids and a pair of keto acids, catalysed by a group of enzymes called **transaminases** (recently, aminotransferases).

Transamination, quantitatively, is the most important reaction in the amino acid metabolism. The salient features of transamination are:
1. All transaminases require **pyridoxal phosphate (PLP)**, a coenzyme derived from vitamin B_6.
2. Specific transaminases exist for each pair of amino and keto acids. However, only two-namely, aspartate transaminase and alanine transaminase make a significant contribution for transamination.
3. There is no free NH_3 liberated; only the transfer of amino group occurs.
4. Transamination is **reversible**.
5. In the cells, all the amino acids are not readily available in a proportion needed for protein biosynthesis. Transamination is very important for the redistribution of amino groups and **production of non-essential amino acids**, as per the requirement of the cell.
6. It involves both catabolism (degradation) and anabolism (synthesis) of amino acids. Transamination is ultimately responsible for the synthesis of non-essential amino acids.
7. Transamination diverts the excess amino acids towards **energy generation**.
8. The amino acids undergo transamination to finally concentrate nitrogen in glutamate. Glutamate is the only amino acid that undergoes oxidative deamination to a significant extent to liberate free NH_3 for urea synthesis.
9. All amino acids except lysine, threonine, proline and hydroxyproline participate in transamination.
10. Transamination is not restricted to α-amino groups only. For instance, δ-amino group of ornithine is transaminated.
11. Serum transaminases are important for diagnostic and prognostic purposes. Serum glutamate pyruvate transaminase (SGPT) or alanine transaminase (ALT) is elevated in all liver diseases. Serum glutamate oxaloacetate transaminase (SGOT) or aspartate transaminase (AST) is increased in myocardial infarction.

Mechanism of transamination
Transamination occurs in two stages
1. Transfer of the amino group to the coenzyme pyridoxal phosphate (bound to the coenzyme) to form pyridoxamine phosphate.
2. The amino group of pyridoxamine phosphate is then transferred to a keto acid to produce a new amino acid and the enzyme with PLP is regenerated.



All the transaminases require **pyridoxal phosphate (PLP)**, a derivative of vitamin B_6. The aldehyde group of PLP is linked with ε-amino group of lysine residue, at the active site of the enzyme forming a **Schiff base** (imine linkage). When an amino acid (substrate) comes in contact with the enzyme, it displaces lysine and a new Schiff base linkage is formed. The amino acid-PLP Schiff base tightly binds with the enzyme by non covalent forces. Snell and Braustein proposed a **Ping Pong Bi Bi** mechanism involving a series of intermediates (aldimines and ketimines) in transamination reaction.

DEAMINATION

The removal of amino group from the amino acids as NH_3 is deamination. Transamination (discussed above) involves only the shuffling of amino groups among the amino acids. On the other hand, deamination results in the liberation of ammonia for urea synthesis. Simultaneously, the carbon skeleton of amino acids is converted to keto acids. Deamination may be either oxidative or non-oxidative.

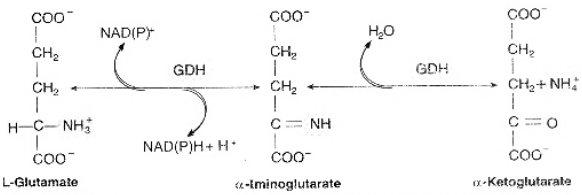
Although transamination and deamination are separately discussed, they occur simultaneously, often involving glutamate as the central molecule. For this reason, some authors use the term **transdeamination** while describing the reactions of transamination and deamination, particularly involving glutamate.

I. Oxidative deamination

Oxidative deamination is the liberation of free ammonia from the amino group of amino acids coupled with oxidation. This takes place mostly in liver and kidney. The purpose of oxidative deamination is to provide NH_3 for urea synthesis and a-keto acids for a variety of reactions, including energy generation.

Role of glutamate dehydrogenase:

In the process of transamination, the amino groups of most amino acids are transferred to α-ketoglutarate to produce glutamate. Thus, glutamate serves as a 'collection centre' for amino groups in the biological system. Glutamate rapidly undergoes oxidative deamination, catalysed by glutamate dehydrogenase (GDH) to liberate ammonia. This enzyme is unique in that it can utilize either NAD^+ or $NADP^+$ as coenzyme. Conversion of glutamate to α-ketoglutarate occurs through the formation of an intermediate, a-iminoglutarate (**Fig**).



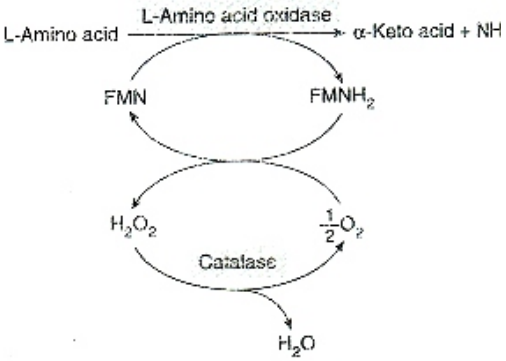
Glutamate dehydrogenase catalysed reaction is important as it reversibly links up glutamate metabolism with TCA cycle through α ketoglutarate. GDH is involved in both catabolic and anabolic reactions.

Regulation of GDH activity: Glutamate dehydrogenase is a zinc containing mitochondrial enzyme. It is a complex enzyme consisting of six identical units with a molecular weight of 56,000 each. GDH is controlled by allosteric regulation. GTP and ATP inhibit whereas GDP and ADP activate-glutamate dehydrogenase. Steroid and thyroid hormones inhibit GDH.

After ingestion of a protein-rich meal, liver glutamate level is elevated. It is converted to α ketoglutarate with liberation of

NH_3. Further, when the cellular energy levels are low, the degradation of glutamate is increased to provide α - ketoglutarate which enters TCA cycle to liberate energy.

Oxidative deamination by amino acid oxidases : L-Amino acid oxidase and D-amino acid oxidase are flavoproteins, possessing FMN and FAD, respectively. They act on the corresponding amino acids (L or D) to produce α-keto acids and NH_3. In this reaction, oxygen is reduced to H_2O_2 which is later decomposed by catalase (Fig.).



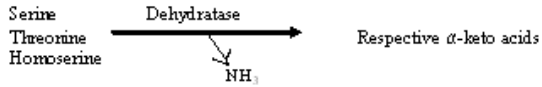
The activity of L-amino acid oxidase is much low while that of D-amino acid oxidase is high in tissues (mostly liver and kidney). L-Amino acid oxidase does not act on glycine and dicarboxylic acids. This enzyme, due to its very low activity, does not appear to play any significant role in the amino acid metabolism.

Fate of D-amino acids: D-Amino acids are found in plants and microorganisms. They are, however, not present in the mammalian proteins. But D-amino acids are regularly taken in the diet and metabolised by the body. D-Amino acid oxidase converts them to the respective a-keto acids by oxidative deamination. The a-keto acids so produced undergo transamination to be converted to L-amino acids which participate in various metabolisms. Keto acids may be oxidized to generate energy or serve as precursor for glucose and fat synthesis. Thus, D-amino acid oxidase is important as it initiates the first step for the conversion of unnatural D-amino acids to L-amino acids in the body.

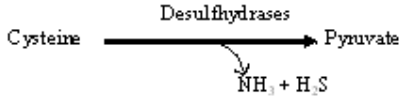
II. Non-oxidative deamination

Some of the amino acids can be deaminated to liberate NH_3 without undergoing oxidation

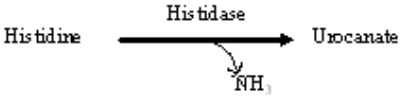
a) Amino acid dehydrases: Serine, threonine and homoserine are the hydroxy amino acids. They undergo non-oxidative deamination catalysed by PLP-dependent dehydrases (dehydratases).



b) Amino acid desulfhydrases: The sulfur amino acids, namely cysteine and homocysteine, undergo deamination coupled with desulfhydration to give keto acids.



c) Deamination of histidine: The enzyme histidase acts on histidine to liberate NH3 by a nonoxidative deamination process.



METABOLISM OF AMMONIA

Ammonia is constantly being liberated in the metabolism of amino acids (mostly) and other nitrogenous compounds. At the physiological pH, **ammonia exists as ammonium (NH_4^+) ion**.

I. Formation of ammonia

The production of NH_3 occurs from the amino acids (transamination and deamination), biogenic amines, amino group of purines and pyrimidines and by the action of intestinal bacteria (urease) on urea.

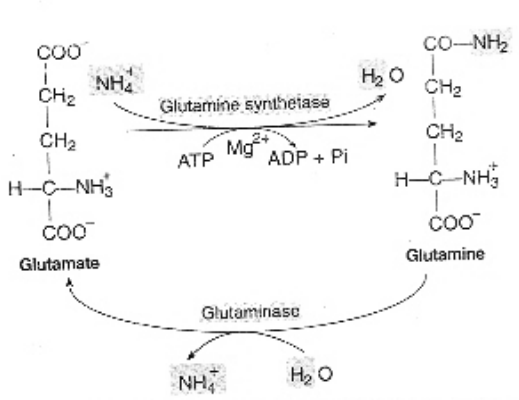
II. Transport and storage of NH_3

Despite a regular and constant production of NH_3 from various tissues, its concentration in the circulation is surprisingly low (normal plasma 10-20 μg/dl). This is mostly because the body has an efficient mechanism for NH3 transport and its immediate utilization for urea synthesis. The transport of ammonia mostly occurs in the form of glutamine or alanine and not as free ammonia. Alanine is important for NH_3 transport from muscle to liver by glucose-alanine cycle.

Role of glutamine: Glutamine is a storehouse of NH_3, It is present at the highest concentration (8 mg/dl in adults) in blood among the amino acids. Glutamine serves as a storage and transport form of NH_3. Its synthesis mostly occurs in liver, brain and muscle. Ammonia is removed from the brain predominantly as glutamine. Glutamine is freely diffusible in tissues, hence easily transported.

Glutamine synthetase (a mitochondrial enzyme) is responsible for the synthesis of glutamine from glutamate and ammonia. This reaction is unidirectional and requires ATP and Mg^{2+} ions.

Glutamine can be deaminated by hydrolysis to release ammonia by glutaminase (fig.) an enzyme mostly found in kidney and intestinal cells.



III. Functions of ammonia

Ammonia is not just a waste product of nitrogen metabolism. It is involved (directly or via glutamine) for the synthesis of many compounds in the body. These include non-essential amino acids, purines, pyrimidines, amino sugars, asparagine etc. Ammonium ions (NH_4^+) are very important to maintain acid-base balance of the body. .

IV. Disposal of ammonia

The organisms, during the course of evolution, have developed different mechanisms for the disposal of ammonia from the body. The animals in this regard are of three different types

a) Ammoniotelic: The aquatic animals dispose off NH_3 into the surrounding water.

b) Uricotelic: Ammonia is converted mostly to uric acid e.g. reptiles and birds.

c) Ureotelic: The mammals including man convert NH_3 to urea. Urea is a non-toxic and soluble compound, hence easily excreted.

v. Toxicity of ammonia

Even a marginal elevation in the blood ammonia concentration is harmful to the brain. Ammonia, when it accumulates in the body, results in slurring of speech and blurring of the vision and causes tremors. It may lead to coma and, finally, death, if not corrected.

Hyper-ammonemia: Elevation in blood NH3 level may be genetic or acquired. Impairment in urea synthesis due to a defect in anyone of the five enzymes is described in urea synthesis. All these disorders lead to hyperammonemia and cause mental retardation. The acquired hyperammonemia may be due to hepatitis, alcoholism etc. where the urea synthesis becomes defective, hence NH_3 accumulates.

Explanation for NH_3 toxicity: The reaction catalysed by glutamate dehydrogenase probably explains the toxic affects of NH_3 in brain.

Accumulation of NH_3 shifts the equilibrium to the right with more glutamate formation, hence more utilization of α-ketoglutarate. α-Ketoglutarate is a key intermediate in TCA cycle and its depleted levels impair the TCA cycle. The net result is that production of energy (ATP) by the brain is reduced. The toxic effects of NH_3 on brain are, therefore, due to impairment in ATP formation.

UREA CYCLE:

Urea is the end product of protein metabolism (amino acid metabolism). The nitrogen of amino acids converted to ammonia (as described above) is toxic to the body. It is converted to urea and detoxified. As such, urea accounts for 80-90% of the nitrogen containing substances excreted in urine.

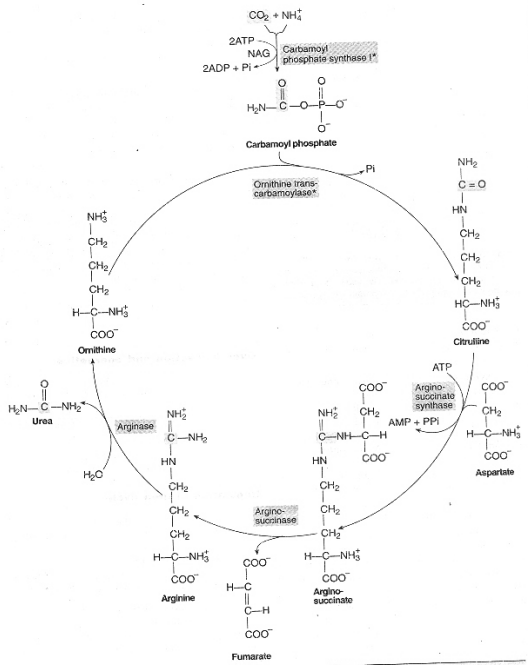
Urea is synthesized in liver, and transported to kidneys for excretion in urine. Urea cycle is the first metabolic cycle that was elucidated by Hans Krebs and Kurt Henseleit (1932), hence it is known as Krebs-Henseleit cycle. The individual reactions, however, were described in more detail later on by Ratner and Cohen.

Urea has two amino (-NH_2) groups, one derived from NH_3 and the other from aspartate. Carbon atom is supplied by CO_2 Urea synthesis is a five-step cyclic process, with five distinct enzymes. The first two enzymes are present in mitochondria while the rest are localized in cytosol. The details of urea cycle are described next (Fig).

1. Synthesis of carbamoyl phosphate: Carbamoyl phosphate synthase I (CPS I) of mitochondria catalyses the condensation of NH: ions with CO_2 to form carbamoyl phosphate. This step consumes two ATP and is irreversible, and rate-limiting. CPS I requires N-acetylglutamate for its activity. Another enzyme, carbamoyl phosphate synthase II (CPS II)-involved in pyrimidine synthesis-is present in cytosol. It accepts amino group from glutamine and does not require N-acetylglutamate for its activity.

2. Formation of citrulline: Citrulline is synthesized from carbamoyl phosphate and ornithine by ornithine transcarbamoylase. Ornithine is regenerated and used in urea cycle. Therefore, its role is comparable to that of oxaloacetate in citric acid cycle. Ornithine and citrulline are basic amino acids. (They are never found in protein structure due to lack of codons). Citrulline produced in this reaction is transported to cytosol by a transporter system.

3. Synthesis of arginosuccinate: Arginosuccinate synthase condenses citrulline with aspartate to produce arginosuccinate. The second amino group of urea is incorporated in this reaction. This step requires ATP which is cleaved to AMP and pyrophosphate (PPi). The latter is immediately broken down to inorganic phosphate (Pi).



4. Cleavage of arginosuccinate: Arginosuccinase cleaves arginosuccinate to give arginine and fumarate. Arginine is the immediate precursor for urea. Fumarate liberated here provides a connecting link with TCA cycle, gluconeogenesis etc.

5. Formation of urea: Arginase is the fifth and final enzyme that cleaves arginine to yield urea and ornithine. Ornithine, so regenerated, enters mitochondria for its reuse in the urea cycle. Arginase is activated by $C0^{2+}$ and Mn^{2+}. Ornithine and lysine compete with arginine (competitive inhibition) Arginase is mostly found in the liver; while the rest of the enzymes (four) of urea cycle are also present in other tissues. For this reason, arginine synthesis may occur to varying degrees in many tissues. But only the liver can ultimately produce urea.

Overall reaction and energetics

The urea cycle is irreversible and consumes 4 ATP. Two ATP are utilized for the synthesis of carbamoyl phosphate. One ATP is converted to AMP and PPi to produce arginosuccinate which equals 2 ATP. Hence 4 ATP are actually consumed.

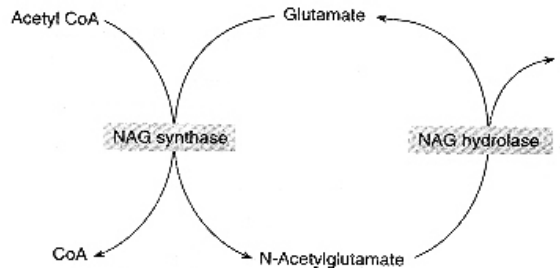
$NH_4^+ + CO_2 +$ Aspartate + 3ATP → Urea + Fumarate + 2 ADP + 2 Pi + AMP + PPi

Regulation of urea cycle

The first reaction catalysed by carbamoyl phosphate synthase I (CPS I) is rate-limiting reaction or committed step

in urea synthesis. CPS I is allosterically activated by N-acetylglutamate (NAG). It is synthesized from glutamate and acetyl CoA by synthase and degraded by a hydrolase.

The rate of urea synthesis in liver is correlated with the concentration of N-acetylglutamate. The consumption of a protein-rich meal increases the level of NAG in liver, leading to enhanced urea synthesis.



Carbamoyl phosphate synthase I and glutamate dehydrogenase are localized in the mitochondria. They coordinate with each other in the formation of NH_3 and its utilization for the synthesis of carbamoyl phosphate. The remaining four enzymes of urea cycle are mostly controlled by the concentration of their respective substrates.

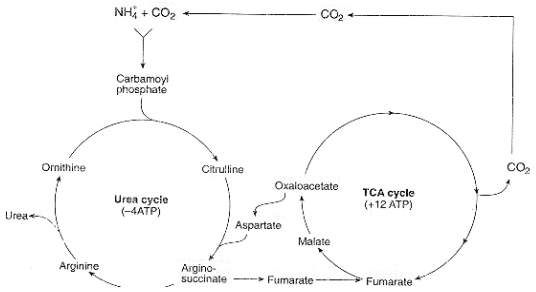
Disposal of urea

Urea produced in the liver freely diffuses and is transported in blood to kidneys and excreted. A small amount of urea enters the intestine where it is broken down to CO_2 and NH_3 by the bacterial enzyme urease. This ammonia is either lost in the feces or absorbed into the blood. In renal failure, the blood urea level is elevated (uremia), resulting in diffusion of more urea into intestine and its breakdown to NH_3, Hyperammonemia (increased blood NH_3) is commonly seen in patients of kidney failure. For these patients, oral administration of antibiotics (neomycin) to kill intestinal bacteria is advised.

Integration between urea cycle and TCA cycle

Urea cycle is linked with TCA cycle in three different ways (Fig.)

1. The production of fumarate in urea cycle is the most important integrating point with TCA cycle. Fumarate is converted to malate and then to oxaloacetate in TCA cycle. Oxaloacetate undergoes transamination to produce aspflrtate which enters urea cycle. Here, it combines with citrulline to produce arginosuccinate. Oxaloacetate is an important metabolite which can combine with acetyl CoA to form citrate and get finally oxidized. Oxaloacetate can also serve as a precursor for the synthesis of glucose (gluconeogenesis).



2. ATP (12) are generated in the TCA cycle while ATP (4) are utilized for urea synthesis.

3. Citric acid cycle is an important metabolic pathway for the complete oxidation of various metabolites to C02 and H_2O. The CO_2 liberated in TCA cycle (in the mitochondria) can be utilized in urea cycle.

Metabolic disorders of urea cycle

Metabolic defects associated with each of the five enzymes of urea cycle have been reported (Table 15.1). All the disorders invariably lead to a build-up in blood ammonia (hyperammonemia), leading to toxicity. Other metabolites of urea cycle also accumulate which, however, depends on the specific enzyme defect. The clinicCll symptoms associated with defect in urea cycle enzymes include vomiting, lethargy, irritability, ataxia and mental retardation.

J7. INTEGRATION OF METABOLISM OF CARBOHYDRATES, LIPIDS AND PROTEINS

Though metabolism of each of major food nutrients, viz., carbohydrates, lipids and proteins have been considered separately for the sake of convenience, it actually takes place simultaneously in the intact animal and are closely interrelated to one another. The metabolic processes involving these three major food nutrients and their interrelationship can be broadly divided into three stages (**Fig.**):

I Stage: Stage of hydrolysis to simpler units
II Stage: Preparatory stage
III Stage: Oxidative stage-Aerobic final (TCA Cycle).

I Stage: Stage of Hydrolysis to Simpler Units:

- The complex polysaccharides, starch/ glycogen are broken down to glucose; and disaccharides are hydrolyzed to monosaccharides in GI tract by various carbohydrate-splitting enzymes present in digestive juices.
- Similarly, principal lipids, triacylglycerol (TAG) is hydrolyzed to form Free Fatty Acids and glycerol.
- Proteins are hydrolyzed by proteolytic enzymes to amino acids.

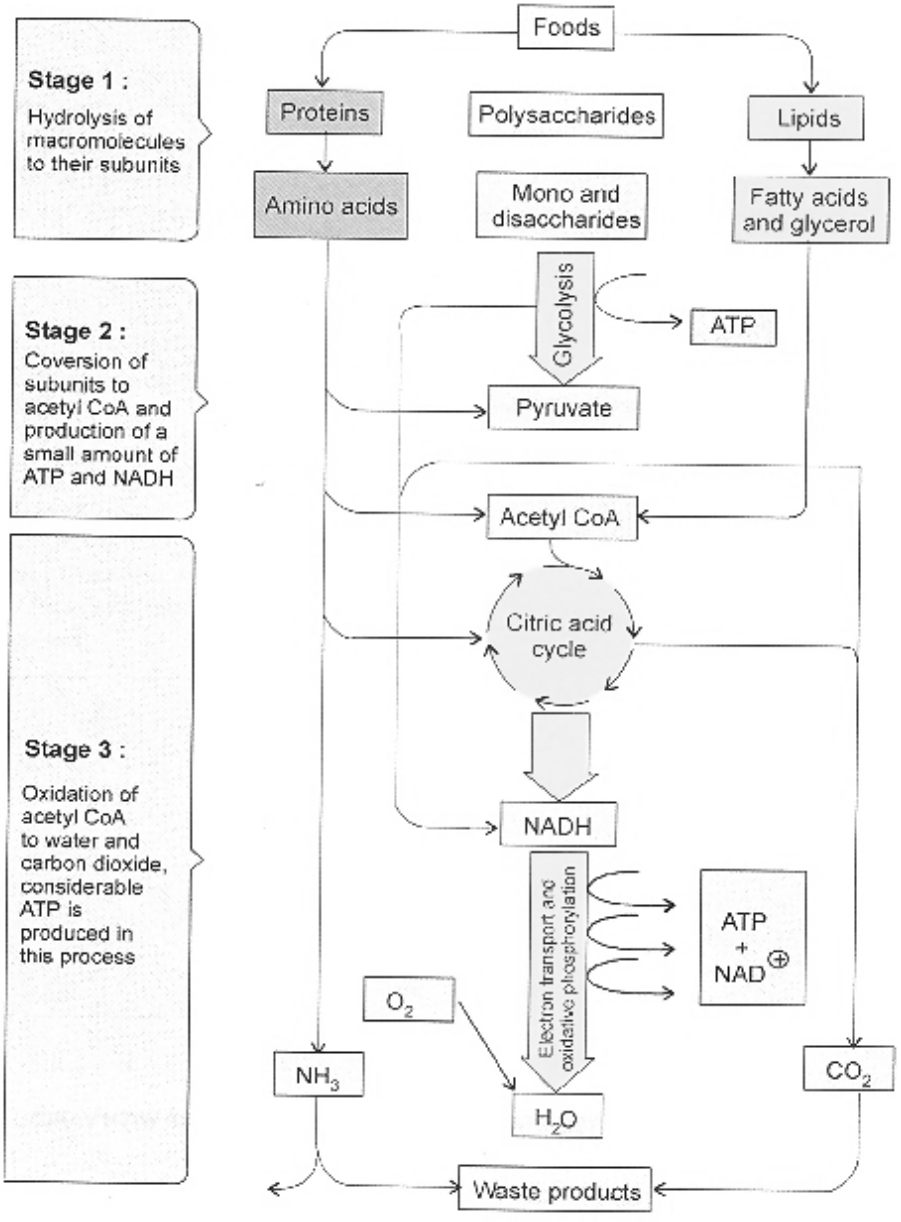
The above is the prelude to either further synthesis new substances or for their oxidation. ***Very little of energy is produced in this hydrolytic phase and it is dissipated away as heat. There is no storage of energy at s stage.***

II Stage: Preparatory Stage:

- The monosaccharide glucose runs through the glycolytic reactions to produce the 3-C keto acid pyruvic acid (PA) in the cytosol, which in turn is transported to mitochondrion where it undergoes oxidative decarboxylation to produce 2-C compound **"acetyl CoA" (" active"** acetate).
- The glycerol of fat, either goes into formation of glucose (gluconeogenesis) or by entering the same glycolytic pathway through the triose-Phosphate, forms Pyruvic Acid and then finally 2-C compound "acetyl-CoA"
- The fatty acids undergo principally β-oxidation and form several molecules of "acetyl-CoA".
- The amino acids are deaminated/and/or transaminated first and the C-skeleton is metabolized differently from amino acid to amino acid

- In the case of amino acids viz., Glycine, Alanine, Serine, Cysteine/ Cystine and threonine when catabolized form pyruvic acid (PA) similar to carbohydrates and is finally converted to 'Acetyl-CoA'
- In the case of amino acids, viz. Glutamic acid, Histidine, Proline and Hydroxy-proline, Arginine and Ornithine produces a-ketoglutaric acid when catabolized and thus they enter the TCA cycle.
- Yet a few others like Leucine, Phenyl alanine, Tyrosine and Isoleucine yield acetate or acetoacetate, the latter can be converted to "acetyl CoA".

During the second stage (glycolysis, β-oxidation, etc.) relatively small amount of energy is produced and this is stored as ATP.

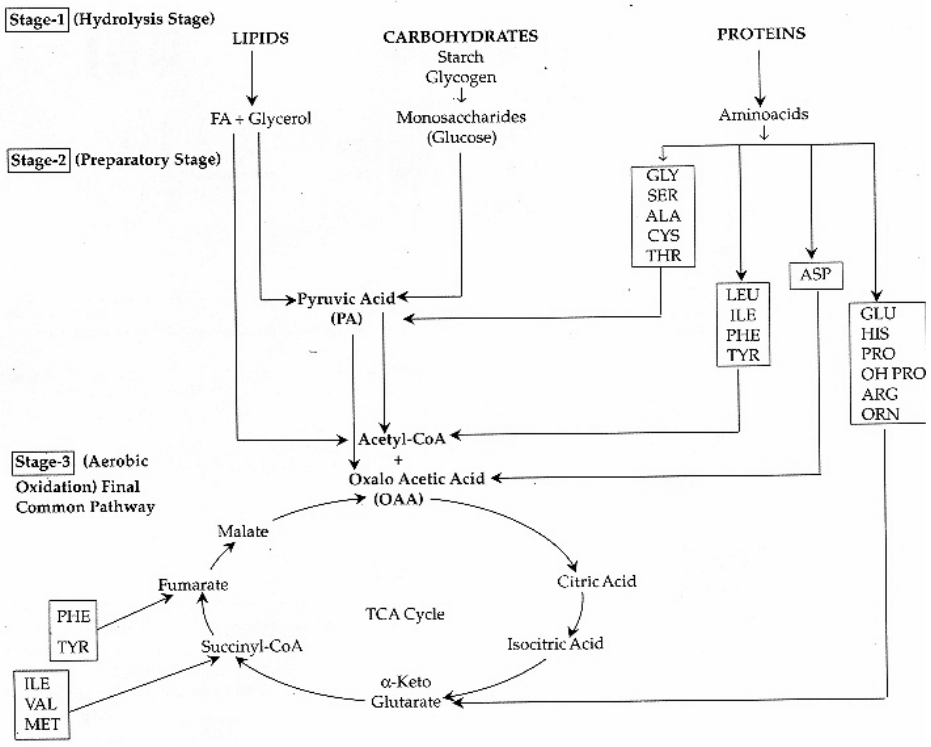


II Stage: Oxidative Stage: Aerobic Final (TCA Cycle)

In presence of oxygen, acetyl-CoA is oxidized to CO_2 and H_2O by common final pathway TCA cycle.

The carbohydrates, lipids and proteins all form acetate or some other intermediates like oxaloacetate (OAA), α-Ketoglutarate, succinyl-CoA, or fumarate, which are all intermediates of TCA cycle. Having gained entry into the TCA cycle at any site, two of carbons of "**citrate**" constituting an acetate moiety are oxidized finally to CO_2 and H_2O and the energy of oxidation by the electron transport chain is captured as energy-rich PO_4-ATP mostly. **This stage yields the largest amount of energy** of all three stages. Thus, the pathways are similar to a large extent and identical in the final stage of oxidation of the metabolites, whether derived from carbohydrates, lipids or proteins.

This is schematically represented in **Figure,** along with the entry of various amino acids.



Interconversion between the Three Principal Components

I. Carbohydrates

1. Carbohydrates can form lipids: Through formation of: (a) α-glycero-Phosphate from glycerol or di-hydroxy acetone Phosphate (from glycolysis) which is necessary for Triacyl glycerol (TAG) and (b) Fatty Acids from acetyl-CoA-extra-mitochondrial *de novo* synthesis.
2. Carbohydrates can form non-essential amino acids: Through amination of α-keto acids, viz., pyruvic acid (PA), oxalo-acetic acid (OAA) and α-ketoglutarate to form amino acids alanine, aspartate and glutamate respectively.

II. Fats:
1. Fatty acids can be converted to some amino acids by forming the dicarboxylic acids like malic acid, oxalo acetic acids and α-ketoglutarate.
2. Fatty acid carbon may theoretically be incorporated into carbohydrates by the acetate running through TCA cycle. But there is no net gain in carbohydrates, since two carbons, equivalent of acetate are oxidized in the cycle.
3. However acetate can form glucose by running through the glyoxylate cycle.
4. Acetone, one of the ketone bodies may be glucogenic. Acetone can be converted to acetol-P which in turn can produce propane diol-P. Propanediol-(P) is glucogenic.

III. Proteins:
Proteins can form both carbohydrates and lipids through the glucogenic and ketogenic amino acids.

Regulation and Control of the Reactions
The ratio of ATP / AMP of the cells/ or tissues seems to decide the extent of its aerobic metabolism.
(a) Inhibition: If the ratio is high (low AMP or ADP level), this will have certain inhibitory effects of certain enzymes of glycolytic- TCA cycle.

1. A high level of ATP and low level of AMP will inhibit the enzyme phosphofructokinase of glycolytic pathway and thereby inhibit glycolysis.
   i. As a result there is accumulation of hexose-P which interacts with UTP to form UDP-G and proceeds to increased glycogen synthesis.
   ii. G-6-P will also be channelized to HMP-shunt leading to increased formation of NADPH which will participate in reductive synthesis, like FA. Synthesis which will be increased.
   iii. The converse happens with low ATP and high AMP levels.
2. Increased ATP / ADP ratio will stimulate PDH-kinase which in turn converts dephosphorylatedactive PDH (pyruvate dehydrogenase complex) to 'inactive' phosphorylated PDH inhibiting the oxidative decarboxylation of pyruvic acid (PA).
3. High ATP / AMP ratio, also lowers the activity of the enzymes Isocitrate dehydrogenase (ICD) of TCA cycle resulting in accumulation of citrate. The oxidation in TCA cycle decreases and ATP production falls.

(b) Stimulation: Increased citric acid levels stimulate the enzyme acetyl-CoA carboxylase. Increased activity of acetyl CoA carboxylase converts acetyl-CoA to malonyl-CoA, the first step in extra-mitochondrial *de novo* Fatty Acid synthesis.

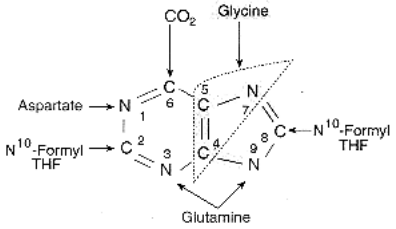
Thus, the acetyl-CoA, in the presence of adequate stores of ATP and low AMP levels, is diverted to the synthesis of fats. The reverse set of conditions operates when the ATP / AMP ratio is low.

J8. METABOLISM OF NUCLEOTIDES

Nucleotides consist of a nitrogenous base, a pentose and a phosphate. The pentose sugar is D-ribose in ribonucleotides of RNA while in deoxyribonucleotides (deoxynucleotides) of DNA, the sugar is 2-deoxy D-ribose. Nucleotides participate in almost all the biochemical processes, either directly or indirectly. They are the structural components of nucleic acids (DNA, RNA), coenzymes, and are involved in the regulation of several metabolic reactions.

BIOSYNTHESIS OF PURINE RIBONUCLEOTIDES

Many compounds contribute to the purine ring of the nucleotides **(Fig.)**.

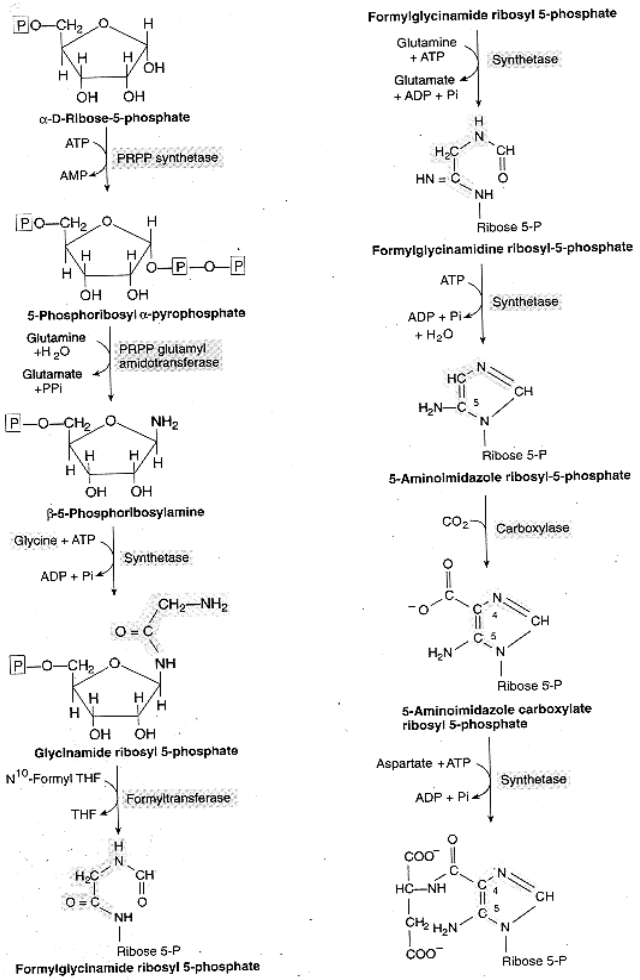


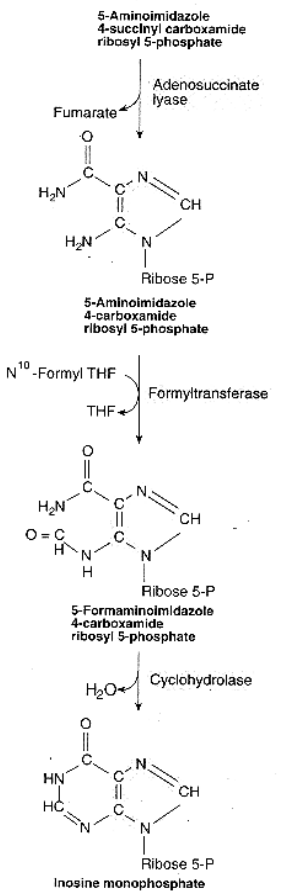
1. N_1 of purine is derived from amino group of aspartate.
2. C_2 and C_8 arise from formate of N^{10}-formyl Tetra Hydro Folate (THF).
3. N_3 and N_9 are obtained from amide group of glutamine.
4. C_4, C_5 and N_7 are contributed by glycine.
5. C_6 directly comes from CO_2

It should be remembered that purine bases are not synthesized as such, but they are formed as ribonucleotides. **The purines are built upon a preexisting ribose 5-phosphate**. Liver is the major site for purine nucleotide synthesis. Erythrocytes, polymorphonuclear leukocytes and brain cannot produce purines.

The pathway for the synthesis of **inosine monophosphate** (IMP or inosinic acid), the 'parent' purine nucleotide is given in **Fig.** The reactions are briefly described below

1. Ribose 5-phosphate, produced in the hexose monophosphate shunt of carbohydrate metabolism is the starting material for purine nucleotide synthesis. It reacts with ATP to form phosphoribosyl pyrophosphate (**PRPP**).
2. Glutamine transfers its amide nitrogen to PRPP to replace pyrophosphate and produce 5-phosphoribosylamine. The enzyme **PRPP glutamyl amidotransferase** is controlled by feedback inhibition of nucleotides (IMP, AMP and GMP). This reaction is the '**committed step**' in purine nucleotide biosynthesis.
3. Phosphoribosylamine reacts with glycine in the presence of ATP to form glycinamide ribosyl 5-phosphate or glycinamide ribotide (GAR).
4. N^{10}-Formyl tetrahydrofolate donates the formyl group and the product formed is formyl glycinamide ribosyl 5-phosphate.
5. Glutamine transfers the second amido amino group to produce formyl glycinamidine ribosyl 5-phosphate.
6. The imidazole ring of the purine is closed in an ATP dependent reaction to yield 5-amino-imidazole ribosyl 5-phosphate.
7. Incorporation of CO_2 (carboxylation) occurs to yield amino-imidazole carboxylate ribosyl 5-phosphate. This reaction does not require the vitamin biotrn and/or ATP which is the case with most of the carboxylation reactions.
8. Aspartate condenses with the product in reaction 7 to form aminoimidazole 4-succinyl carboxamide ribosyl 5-phosphate.
9. Adenosuccinate lyase cleaves off fumarate and only the amino group of aspartate is retained to yield aminoimidazole 4-carboxamide ribosyl 5-phosphate.
10. N^{10}-Formyl tetrahydrofolate donates a o,necarbon moiety to produce formaminoimidazole 4-carboxamide ribosyL 5-phosphate. With this reaction, all the carbon and nitrogen atoms of purine ring are contributed by the respective sources.
11. The final reaction catalysed by cydohydrolase leads to ring closure with an elimination of water molecule. The product obtained is inosine monophosphate (IMP) from which other purine nucleotides can be synthesized.





Inhibitors of purine synthesis

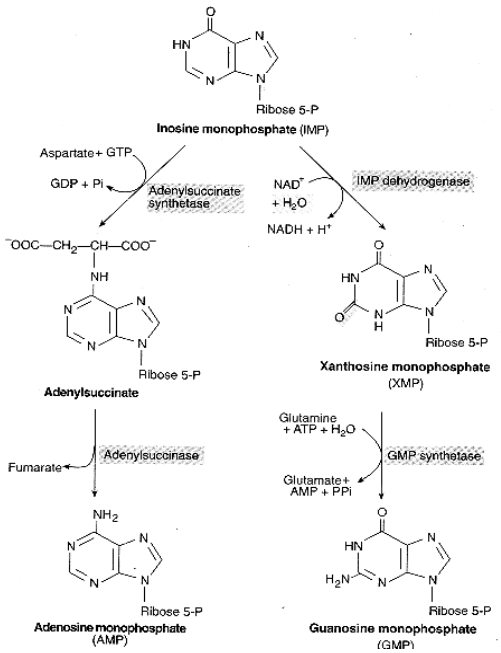
Folic acid (THF) is essential for the synthesis of purine nucleotides (reactions 4 and 10). **Sulfonamides** are the structural analogs of para-aminobenzoic acid (PABA). These sulfa drugs can be used to inhibit the synthesis of folic acid by microorganisms. This indirectly reduces the synthesis of purines and, therefore, the nucleic acids (DNA and RNA). **Sulfonamides** have no influence on humans, since folic acid is not synthesized and is supplied through diet.

The structural analogs of folic acid (e.g. **methotrexate**) are widely used to control cancer. They inhibit the synthesis of purine nucleotides (reaction 4 and 10) and, thus, nucleic acids. Both these reactions are concerned with the transfer of one-carbon moiety (formyl group). These inhibitors also affect the proliferation of normally growing cells. This causes many side-effects including anemia, baldness, scaly skin etc.

Azaserine is an antagonist of glutamine and this can inhibit the reaction 5.

Synthesis of AMP and GMP from IMP:

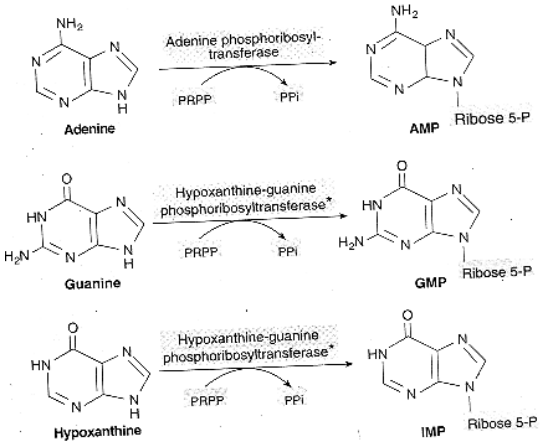
- Inosine monophosphate is the immediate precursor for the formation of AMP and GMP **(Fig.)**. Aspartate condenses with IMP in the presence of GTP to produce adenylsuccinate which, on cleavage, forms AMP.
- For the synthesis of GMP, IMP undergoes NAD+ dependent dehydrogenation to form xanthosine monophosphate (XMP).
- Glutamine then transfers amide nitrogen to XMP to produce GMP.
- 6-Mercaptopurine is an inhibitor of the synthesis of AMP and GMP. It acts on the enzyme adenylsuccinase (of AMP pathway) and IMP dehydrogenase (of GMP pathway).



Formation of purine nucleoside diphosphates and triphosphates:

The nucleoside monophosphates (AMP and GMP) have to be converted to the corresponding di-and triphosphates to participate in most of the metabolic reactions. This is achieved by the transfer of phosphate group from ATP, catalysed by nucleoside monophosphate (NMP) kinases and nucleoside diphosphate (NDP) kinases.

Salvage pathway for Purines



The free purines (adenine, guanine and hypoxanthine) are formed in the normal turnover of nucleic acids (particularly RNA) and also obtained from the dietary sources. The purines can be directly converted to the corresponding

nucleotides and this process is known as '**salvage pathway**' (**Fig.**).

Adenine phosphoribosyl transferase catalyses the formation of AMP from adenine. Hypoxanthineguanine phosphoribosyl transferase (HGPRT) converts guanine and hypoxanthine, respectively, to GMP and IMP. Phosphoribosyl pyrophosphate (PRPP) is the donor of ribose 5-phosphate in the salvage pathway.

The salvage pathway is particularly important in certain tissues such as erythrocytes and brain where de novo (a new) synthesis of purine nucleotides is not operative.

A defect in the enzyme HGPRT causes **Lesch Nyhan syndrome**.

Regulation of purine nucleotide biosynthesis

The purine nucleotide synthesis is well coordinated to meet the cellular demands. The intracellular concentration of **PRPP** regulates purine synthesis to a large extent. This, in turn, is dependent on the availability of ribose 5-phosphate and the enzyme PRPP synthetase.

PRPP glutamyl amido transferase is controlled by a **feedback mechanism** by purine nucleotides. That is, if AMP and GMP are available in adequate amounts to meet the cellular requirements, their synthesis is turned off at the amidotransferase reaction.

Another important stage of regulation is in the conversion of IMP to AMP and GMP. AMP inhibits adenylsuccinate synthetase while GMP inhibits IMP dehydrogenase. Thus, AMP and GMP control their respective synthesis from IMP by a feedback mechanism.

Conversion of ribonucleotides to deoxyribonucleotides

The synthesis of purine and pyrimidine deoxyribonucleotides occurs from ribonucleotides by a reduction at the C_2 of ribose moiety (Fig.). This reaction is catalysed by a multisubunit (two B_1 and two B_2 subunits) enzyme, **ribonucleotide reductase**.

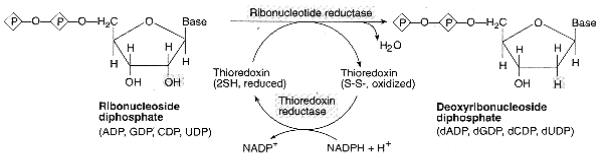


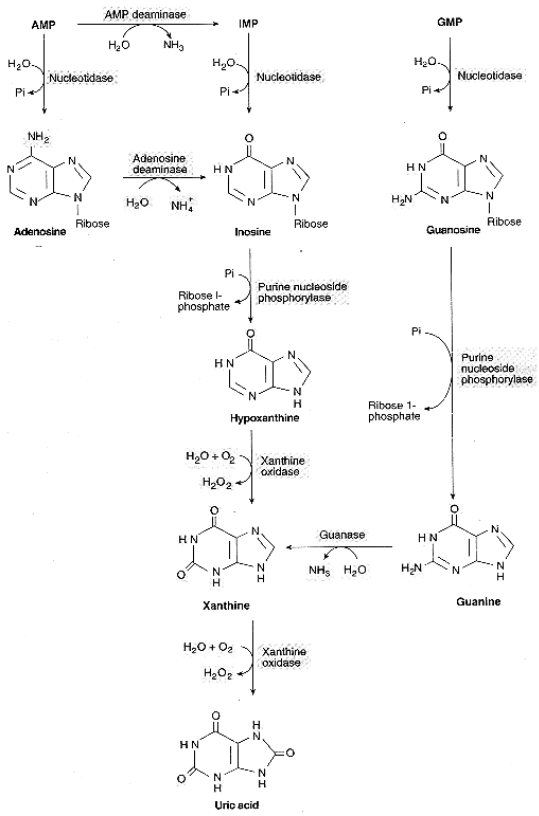
Fig. : Formation of deoxyribonucleotides from ribonucleotides.

Supply of reducing equivalents: The enzyme ribonucleotide reductase itself provides the hydrogen atoms needed for reduction from its sulfhydryl groups. The reducing equivalents, in turn, are supplied by **thioredoxin**, a monomeric protein with two cysteine residues.

NADPH-dependent thioredoxin reductase converts the oxidized thioredoxin to reduced form which can be recycled again and again. **Thioredoxin** thus serves as a **protein cofactor** in an enzymatic reaction.

Regulation of deoxyribonucleotide synthesis: Deoxyribonucleotides are mostly required for the synthesis of DNA. The activity of the enzyme ribonucleotide reductase maintains the adequate supply of deoxyribonudeotides.

Ribonucleotide reductase is a complex enzyme with multiple sites (active site and allosteric sites) that control the formation of deoxyribonucleotides.



DEGRADATION OF PURINE NUCLEOTIDES

The end product of purine metabolism in humans is uric acid. The sequence of reactions in purine nucleotide degradation is given in **Fig.**

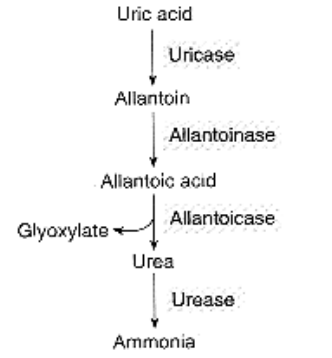


***Fig.:** Degradation of uric acid in animals other than man.*

1. The nucleotide monophosphates (AMP, IMP and GMP) are converted to their respective nucleoside forms (adenosine, inosine and guanosine) by the action of nucleotidase.
2. The amino group, either from AMP or adenosine, can be removed to produce IMP or inosine, respectively.
3. Inosine and guanosine are, respectively, converted to hypoxanthine and guanine (purine bases) by purine nucleoside phosphorylase.

Adenosine is not degraded by this enzyme, hence it has to be converted to inosine.

4. Guanine undergoes deamination by guanase to form **xanthine**.
5. **Xanthine oxidase** is an important enzyme that converts hypoxanthine to xanthine, and xanthine to uric acid. This enzyme contains FAD, molybdenum and iron, and is exclusively found in liver and small intestine. Xanthine oxidase liberates H_2O_2 which is harmful to the tissues. Catalase cleaves H_2O_2 to H_2O and O_2

Uric acid (2, 6, 8-trioxypurine) is the final excretory product of purine metabolism in humans. Uric acid can serve as an important antioxidant by getting itself converted (non-enzymatically) to allantoin. It is believed that the antioxidant role of ascorbic acid in primates is replaced by uric acid, since these animals have lost the ability to synthesize ascorbic acid.

Most animals (other than primates) however, oxidize uric acid by the enzyme uricase to allantoin, where the purine ring is cleaved. Allantoin is then converted to allantoic acid and excreted in some fishes. Further degradation of allantoic acid may occur to produce urea (in amphibians, most fishes and some molluscs) and, later, to ammonia (in marine invertebrates).

DISORDERS OF PURINE METABOLISM

Hrperuricemia and gout: Hyperuricemia refers to an elevation in the serum uric acid concentration. This is sometimes associated with increased uric acid excretion (uricosuria). Gout is a metabolic disease associated with overproduction of uric acid. At the physiological pH, uric acid is found in a more soluble form as sodium urate. In severe hyperuricemia, crystals of sodium urate get deposited in the soft tissues, particularly in the joints. Such deposits are commonly known as **tophi**. This causes inflammation in the joints resulting in a painful gouty **arthritis**. Sodium urate and/or uric acid may also precipitate in kidneys and ureters that results in renal damage and stone formation.

Treatment of gout: The drug of choice for the treatment of primary gout is allopurinol. This is a structural analog of hypoxanthine that competitively inhibits the enzyme xanthine oxidase. Besides the drug therapy, restriction in dietary intake of purines and alcohol is advised. Consumption of plenty of water will also be useful.
The anti-inflammatory drug colchicine is used for the treatment of gouty arthritis. Other anti-inflammatory drugs-such as phenylbutazone, indomethacin, oxyphenbutazone, corticosteroidsare also useful.

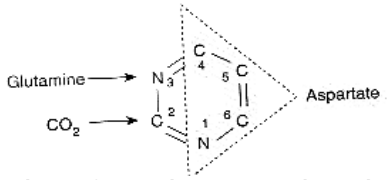
Lesch-Nyhan syndrome: This disorder is due to the deficiency of hypoxanthine-guanine phosphoribosyltransferase (HGPRT), an enzyme of purine salvage pathway (**Fig..5**). Lesch-Nyhan syndrome is a sex-linked metabolic disorder since the structural gene for HGPRT is located on the X-chromosome. It affects only the males and is characterized by excessive uric acid production (often gouty arthritis), and neurological abnormalities such as mental retardation, aggressive behavior, learning disability etc. The patients of this disorder have an irresistible urge to bite their fingers and lips, often causing self mutilation.

Immunodeficiency diseases associated with purine metabolism

Two different immunodeficiency disorders associated with the degradation of purine nucleosides are identified. The enzyme defects are adenosine deaminase and purine nucleoside phosphorylase, involved in uric acid synthesis. The deficiency of **adenosine deaminase (ADA)** causes **severe combined immunodeficiency (SCID)** involving T-cell and usually B-cell dysfunction. It is explained that ADA deficiency results in the accumulation of dATP which is an inhibitor of ribonucleotide reductase and, therefore, DNA synthesis and cell replication.

The deficiency of purine nucleotide phosphorylase is associated with impairment of T-cell .function but has no effect on B-cell function. Uric acid synthesis is decreased and the tissue levels of purine nucleosides and nucleotides are higher. It is believed that dGTP inhibits the development of normal T-cells.

BIOSYNTHESIS OF PYRIMIDINE RIBONUCLEOTIDES



The synthesis of pyrimidines is a much simpler process compared to that of purines. Aspartate, glutamine (amide group) and CO_2 contribute to atoms in the formation of pyrimidine ring (Fig). Pyrimidine ring is first synthesized and then attached to ribose 5-phosphate. This is in contrast to purine nucleotide synthesis wherein purine ring is built upon a pre-existing ribose 5-phosphate. The pathway of pyrimidine synthesis is depicted in Fig., and the salient features are described below.

- Glutamine transfers its amido nitrogen to CO_2 produce carbamoyl phosphate. This reaction ATP-dependent and is catalysed by cytosomal enzyme carbamoyl phosphate synthetase II (CPSII).
- CPS II is activated by ATP and PRPP a inhibited by UTP. Carbamoyl phosphate synthetase I (CPS I) is a mitochondrial enzyme which synthesizes carbamoyl phosphate from ammonia and CO_2 and, in turn urea. Prokaryotes have only one carbamoyl phosphate synthetase which is responsible for the biosynthesis of arginine and pyrimidines.
- Carbamoyl phosphate condenses with aspartate to form carbamoyl aspartate. This reaction catalysed by aspartate transcarbamoylase. Dihydroorotase.catalyses the pyrimidine ring closure with a loss of H_2O.
- The three enzymes-CPS II, aspartate trans-carbamoylase and dihydroorotase are the domains (functional units) of the same protein. This is a good example of a **multifunctional enzyme**.
- The next step in pyrimidine synthesis is an NAD^+ dependent dehydrogenation, leading to the formation of orotate.
- Ribose 5-phosphate is now added to orotate to produce orotidine monophosphate (OMP). This reaction is catalysed by orotate phosphoribosyl-transferase, are enzyme comparable with HGPRT in its function. OMP undergoes decarboxylation to uridine monophosphate (UMP).
- Orotate phosphoribosyltransferase and OMP decarboxylase are **domains** of a single protein. A

defect in this **bifunctional enzyme** causes orotic aciduria.

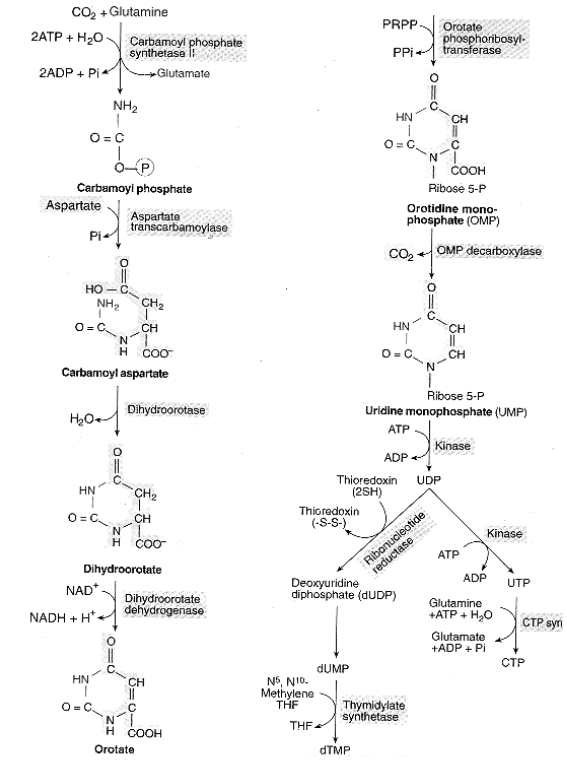


Fig. contd. next column

Fig.: Metabolic pathway for the synthesis of pyrimidine nucleotides.

- By an ATP-dependent kinase reaction, UMP is converted to UDP which serves as a precursor for the synthesis of dUMP, dTMP, UTP and CTP.
- Ribonucleotide reductase converts UDP to dUDP by a thioredoxin-dependent reaction. Thymidylate synthetase catalyses the transfer of a methyl group from NS, NlO-methylene tetrahydrofolate to produce deoxythymidine monophosphate (dTMP).
- UDP undergoes .an ATP-dependent kinase reaction to produce UTP. Cytidine triphosphate (CTP) is synthesized from UTP by amination. CTP synthetase is the enzyme and glutamine provides the nitrogen.

Regulation of pyrimidine synthesis

In bacteria, **aspartate transcarbamoylase (ATCase)** catalyses a committed step in pyrimidine biosynthesis. ATCase is a good example of an enzyme controlled by feedback mechanism by the end product CTP. In certain bacteria, UTP also inhibits ATCase. ATP, however, stimulates ATCase activity.

Carbamoyl phosphate synthetase II (CPS II) is the regulatory enzyme of pyrimidine synthesis in animals. It is activated by PRPP and ATP and inhibited by UDP and UTP. OMP decarboxylase, inhibited by UMP and CMP, also controls pyrimidine formation.

Degradation of pyrimidine nucleotides

The pyrimidine nucleotides undergo similar reactions (dephosphorylation, deamination and cleavage of glycosidic bond) like that of purine nucleotides to liberate the nitrogenous bases cytosine, uracil and thymine. The bases are then degraded to highly soluble products-β-alanine and β-aminoisobutyrate. These are the amino acids which undergo transamination and other reactions to finally produce acetyl CoA and succinyl CoA.

Salvage pathway

The pyrimidines (like purines) can also serve as precursors in the salvage pathway to be converted to the respective nucleotides. This reaction is catalysed by pyrimidine phosphoribosyltransferase which utilizes PRPP as the source of ribose 5-phosphate.

J9. Metabolism of Vitamins

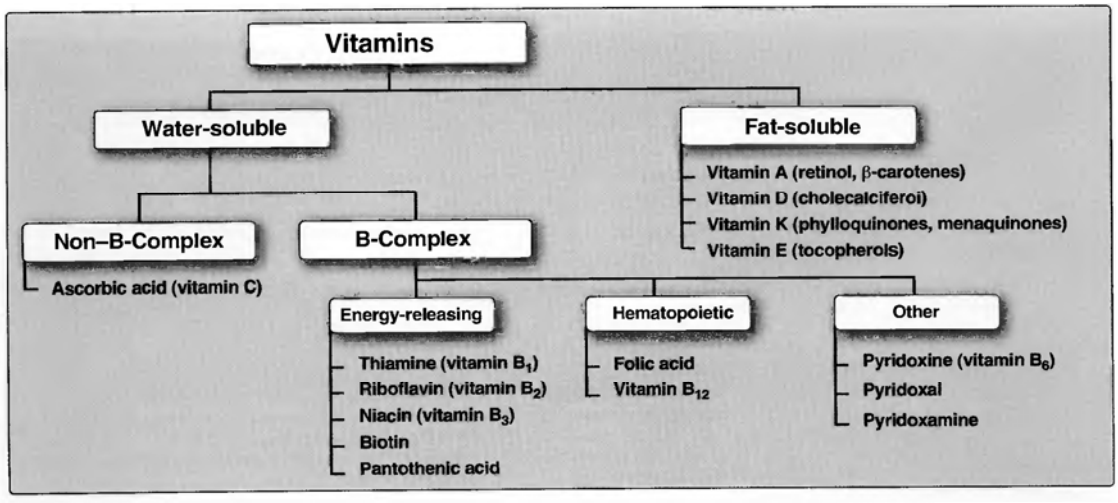


Figure 1 Classification of the vitamins.

Vitamins are chemically unrelated organic compounds that cannot be synthesized by humans and, therefore, must must be supplied by the diet. Nine vitamins (folic acid, cobalamin, ascorbic acid, pyridoxine, thiamine, niacin, riboflavin, biotin, and pantothenic acid) are classified as water-soluble, whereas four vitamins (vitamins A, D, K, and E) are termed fat-soluble (Figure 1). Vitamins are required to perform specific cellular functions, for example, many of the water-soluble vitamins are precursors of coenzymes for the enzymes of intermediary metabolism. In contrast to the water-soluble vitamins, only one fat soluble vitamin (vitamin K) has a coenzyme function. These vitamins are released, absorbed, and transported with the fat of the diet. They are not readily excreted in the urine, and significant quantities are stored in the liver and adipose tissue. In fact, consumption of vitamins A and D in excess of the recommended dietary allowances can lead to accumulation of toxic quantities of these compounds.

Folic acid (or folate), which plays a key role in one-carbon metabolism, is essential for the biosynthesis of several compounds. Folic acid deficiency is probably the most common vitamin deficiency in the United States, particularly among pregnant women and alcoholics.

A. Function of folic acid

Tetrahydrofolate receives one-carbon fragments from donors such as serine, glycine, and histidine and transfers them to intermediates in the synthesis of amino acids, purines, and thymine--a pyrimidine found in DNA.

B. Nutritional anemias

Anemia is a condition in which the blood has a lower than normal concentration of hemoglobin, which results in a reduced ability to transport oxygen. Nutritional anemias--those caused by inadequate intake of one or more essential nutrients---can be classified according to the size of the red blood cells or mean corpuscular volume observed in the individual (Figure 2).

Microcytic anemia, caused by lack of iron, is the most common form of nutritional anemia. The second major category of nutritional anemia results from a deficiency in folic acid or vitamin B12. [Note: These macrocytic anemias are commonly called megaloblastic because a deficiency of folic acid or vitamin B12 causes accumulation of large, immature red precursors, known as megaloblasts, in the the bone marrow.

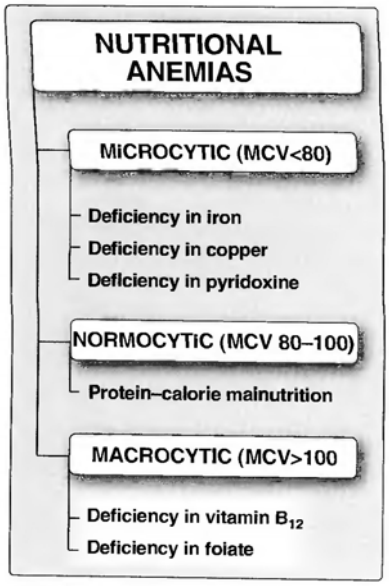


Figure 2
Classification of nutritional anemias by cell size. MCV = Mean corpuscular volume. The normal MCV level for people older than age 18 is between 80 and 100 μm.^{3}

I.

1. Folate and anemia: Inadequate serum levels of folate can be caused by increased demand (for example, pregnancy and lactation), poor absorption caused by pathology of the small intestine, alcoholism, or treatment with drugs that are dihydrofolate reductase inhibitors for example, methotrexate (Figure 3).

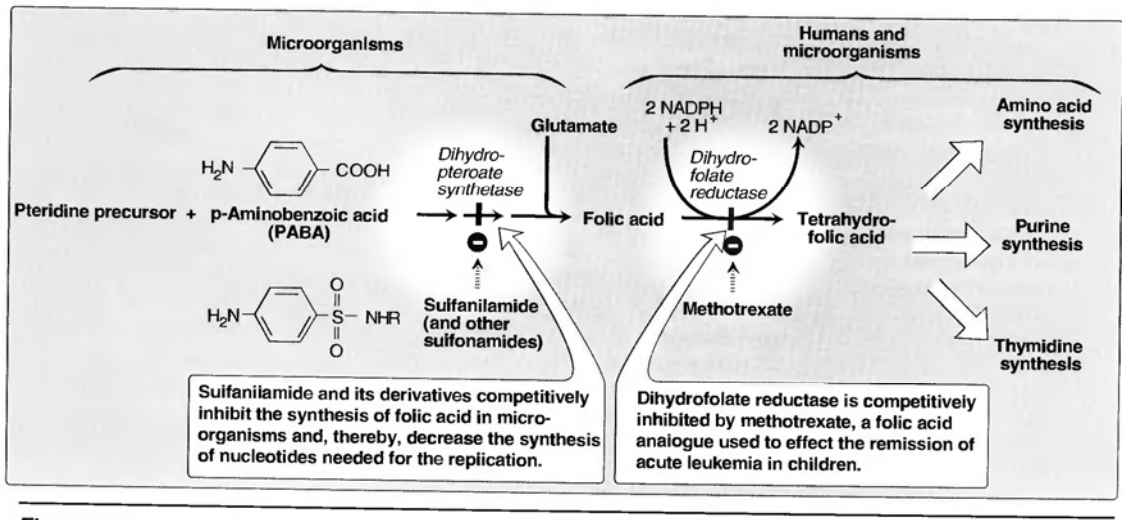


Figure 3
Inhibition of tetrahydrofolate synthesis by sulfonamides and trimethoprim.

A folate-free diet can cause a deficiency within a few weeks. A primary result of folic acid deficiency is megaloblastic anemia, caused by diminished synthesis of purines and thymidine, which leads to an inability of cells to make DNA and, therefore, they cannot divide. [Note: It is important to evaluate the cause of the megaloblastic anemia prior to instituting therapy, because vitamin B12 deficiency indirectly causes symptoms of this disorder.

Folate and neural tube defects in the fetus: Spina bifida and anencephaly, the most common neural tube defects, affect approximately 4000 pregnancies in the United State annually. Folic acid supplementation before conception and during the first trimester has been shown to virtually eliminate the defects. Therefore, all women of childbearing age should consume 0.4 rag/day of folic acid to reduce the risk of having a pregnancy affected by neural tube defects. Adequate folate nutrition must occur at the time of conception because critical folate-dependent development occurs in the first weeks of fetal life--at a time when many women are not yet aware of their pregnancy. The U.S. Food and Drug Administration has authorized the addition of folic acid to enriched grain products, resulting in a dietary supplementation of about 0.1 rag/day. It is estimated that this supplementation will allow approximately fifty percent of all reproductive-aged women to receive 0.4 mg of folate from all sources. However, folic acid intake should not exceed approximately 1 rag/day to avoid complicating the diagnosis of vitamin B12 deficiency.

II. COBALAMIN (VITAMIN B12)

Vitamin B12 is required in humans for two essential enzymatic reactions: the synthesis of methionine and the isomerization of methylmalonyl CoA that is produced during the degradation of some amino acids, and fatty acids with odd numbers of carbon atoms (Figure 5). When the vitamin is deficient, abnormal fatty acids accumulate and become incorporated into cell membranes, including those of the nervous system. This may account for some of the neurologic manifestations of vitamin B12 deficiency.

A. Structure of cobalamin and its coenzyme forms Cobalamin contains a corrin ring system that differs from the porphyrins in that two of the pyrrole rings are linked directly rather than through a methene bridge. Cobalt is held in the center of the corrin ring by four coordination bonds from the nitrogens of the pyrrole groups. The remaining coordination bonds of the cobalt are with the nitrogen of 5,6-dimethylbenzimidazole and with cyanide in commercial preparations of the vitamin in the form of cyanocobalamin (Figure 6). The coenzyme forms of cobalamin are 5'-deoxyadenosyl- cobalamin, in which cyanide is replaced with 5'-deoxyadenosine (forming an unusual carbon-cobalt bond), and methylcobalamin, in which cyanide is replaced by a methyl group (see Figure 6).

B. Distribution of cobalamin

Vitamin B12 is synthesized only by microorganisms; it is not present in plants. Animals obtain the vitamin preformed from their natural bacterial flora or by eating foods derived from other animals. Cobalamin is present in appreciable amounts in liver, whole milk, eggs, oysters, fresh shrimp, pork, and chicken.
C. Folate trap hypothesis

The effects of cobalamin deficiency are most pronounced in rapidly dividing cells, such as the erythropoietic tissue of bone marrow and the mucosal cells of the intestine. Such tissues need both the N5-10-methylene and N10-formyl forms of tetrahydrofolate for the synthesis of nucleotides required for DNA replication. However, in vitamin B12 deficiency, the NS-methyl form of tetrahydrofolate is not efficiently used. Because the methylated form cannot be converted directly to other forms of tetrahydrofolate, the N5-methyl form accumulates, whereas the levels of the other forms decrease. Thus, cobalamin deficiency is hypothesized to lead to a deficiency of the tetrahydrofolate forms needed in purine and thymine synthesis, resulting in the symptoms of megaloblastic anemia.

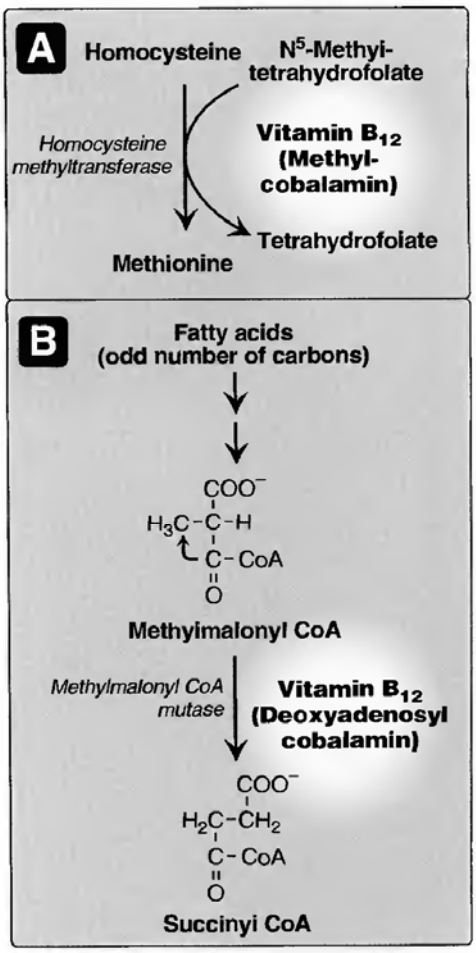


Figure 5 Reactions requiring cofactor forms of vitamin B_{12}.

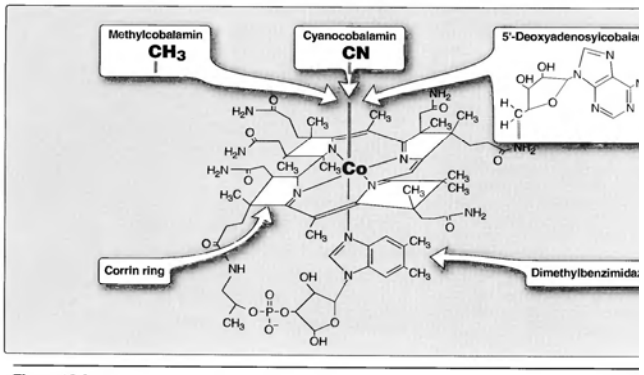


Figure 28.6
Structure of vitamin B_{12} (cyanocobalamin) and its coenzyme forms (methylcobalamin and 5'-deox cobalamin).

D. Clinical indications for vitamin B12

In contrast to other water-soluble vitamins, significant amounts (4 to 5 rag) of vitamin B12 are stored in the body. As a result, it may take several years for the clinical symptoms of B12 deficiency to develop in individuals who have had a partial or total gastrectomy (who, therefore, become intrinsic factor-deficient) and can no longer absorb the vitamin.

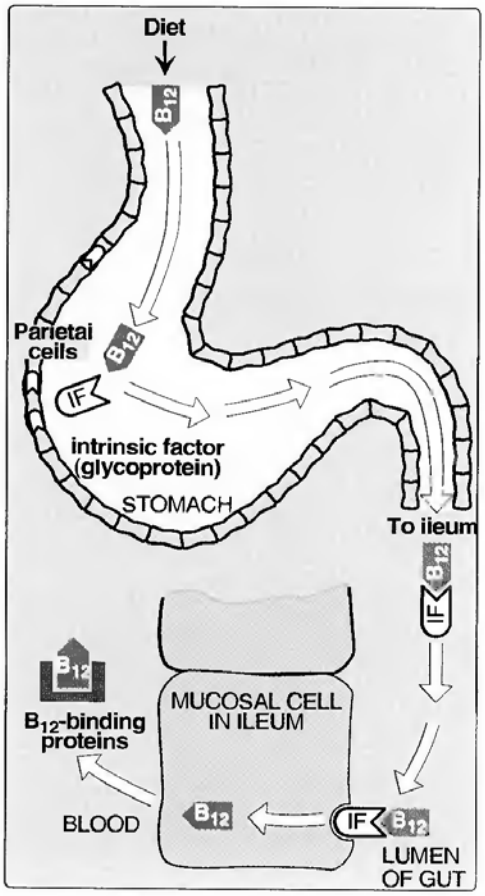


Figure 7 Absorption of vitamin B_{12}.

1. Pernicious anemia:

Vitamin B12 deficiency is rarely a result of an absence of the vitamin in the diet. It is much more common to find deficiencies in patients who fail to absorb the vitamin from the intestine, resulting in pernicious anemia. The disease is most commonly a result of an autoimmune destruction of the gastric parietal cells that are responsible for the synthesis of a glycoprotein called intrinsic factor. Normally, vitamin B12 obtained from the diet binds to intrinsic factor in the intestine (Figure 7). The cobalamin--intrinsic factor complex travels through the gut and eventually binds to specific receptors on the surface of mucosal cells of the ileum. The bound cobalamin is transported into the mucosal cell and, subsequently, into the general circulation, where it is carried by B12-binding proteins. Lack of intrinsic factor prevents the absorption of vitamin B12, resulting in pernicious anemia. Patients with cobalamin deficiency are usually anemic, but later in the development of the disease they show neuropsychiatric symptoms. However, central nervous system (CNS) symptoms may occur in the absence of anemia. The CNS effects are irreversible and occur by mechanisms that appear to be different from those described for megaloblastic anemia. The disease is treated by giving high-dose B12 orally, or intramuscular injection of cyanocobalamin. Therapy must be continued throughout the lives of patients with pernicious anemia. [Note: Folic acid administration alone reverses the hematologic abnormality and, thus, masks the deficiency, which can then proceed to severe neurologic dysfunction and pathology; therefore, megaloblastic anemia should not be treated with folic acid alone, but rather with a combination of folate and vitamin B12.]

III. ASCORBIC ACID (VITAMIN C)

The active form of vitamin C is ascorbate acid (Figure 8). The main function of ascorbate is as a reducing agent in several different reactions. Vitamin C has a well-documented role as a coenzyme in hydroxylation reactions, for example, hydroxylation of prolyl- and lysyl-residues of collagen . Vitamin C is, therefore, required for the maintenance of normal connective tissue, as well as for wound healing. Vitamin C also facilitates the absorption of dietary iron from the intestine.

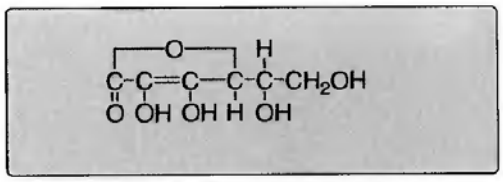


Figure 8 Structure of ascorbic acid.

A. Deficiency of ascorbic acid

A deficiency of ascorbic acid results in scurvy, a disease characterized by sore, spongy gums, loose teeth, fragile blood vessels, swollen joints, and anemia. Many of the deficiency symptoms can be explained by a deficiency in the hydroxylation of collagen, resulting in defective connective tissue.

B. Prevention of chronic disease:

Vitamin C is one of a group of nutrients that includes vitamin E and 13-carotene, which are known as anti-oxidants. Consumption of diets rich in these compounds is associated with a decreased incidence of some chronic diseases, such as coronary heart disease and certain cancers. However, clinical trials involving supplementation with the isolated antioxidants have failed to determine any convincing beneficial effects.

IV. PYRIDOXINE (VITAMIN B6)

Vitamin B6 is a collective term for pyridoxine, pyridoxal, and pyridoxamine, all derivatives of pyridine. They differ only in the nature of the functional group attached to the ring (Figure 10).

Pyridoxine occurs primarily in plants, whereas pyridoxal and pyridoxamine are found in foods obtained from animals. All three compounds can serve as precursors of the biologically active coenzyme, pyridoxal phosphate. Pyridoxal phosphate functions as a coenzyme for a large number of enzymes, particularly those that catalyze reactions involving amino acids.

Reaction type	Example
Transamination	Oxaloacetate + glutamate ⇄ aspartate + α-ketoglutarate
Deamination	Serine → pyruvate + NH_3
Decarboxylation	Histidine → histamine + CO_2
Condensation	Glycine + succinyl CoA → δ-aminolevulinic acid

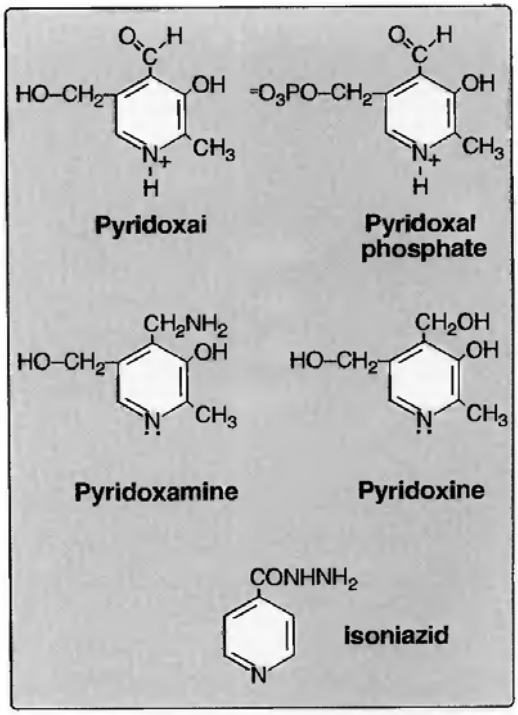


Figure 10
Structures of vitamin B_6 and and the anti-tuberculosis drug isoniazid.

A. Clinical indications for pyridoxine:

Isoniazid (isonicotinic acid hydrazide), a drug frequently used to treat tuberculosis, can induce a B6 deficiency by forming an inactive derivative with pyridoxal phosphate. Dietary supplementation with B6 is, thus, an adjunct to isoniazide treatment. Otherwise, dietary deficiencies in pyridoxine are rare but have been observed in new-born infants fed formulas low in vitamin B6, in women taking oral contraceptives, and in alcoholics.

B. Toxicity of pyridoxine

Neurologic symptoms have been observed at intakes of greater than 2 g/day. Substantial improvement, but not complete recovery, occurs when the vitamin is discontinued.

V. THIAMINE (VITAMIN B1)

Thiamine pyrophosphate (TPP) is the biologically active form of the vitamin, formed by the transfer of a pyrophosphate group from ATP to thiamine (Figure 11).

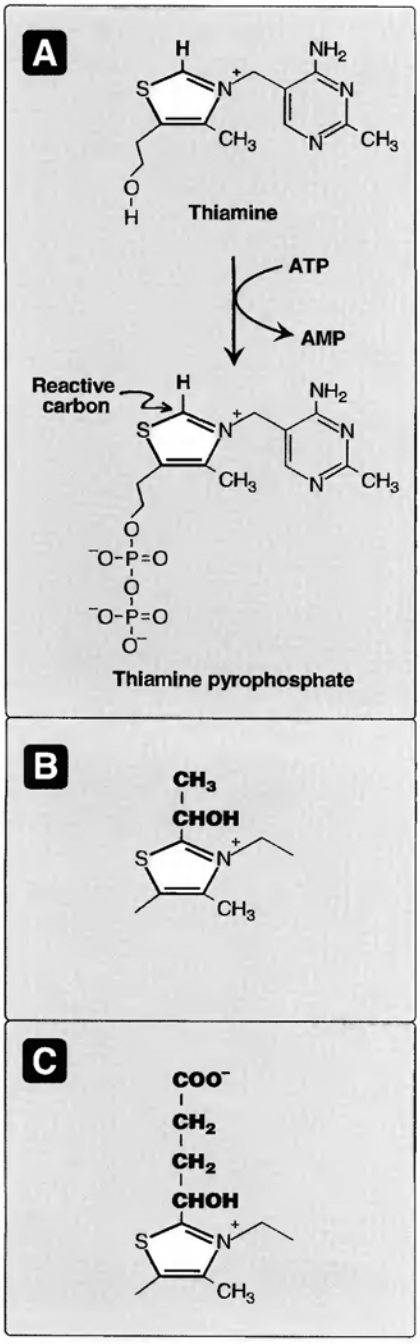


Figure 11
A. Structure of thiamine and its cofactor form, thiamine pyrophosphate. B. Structure of intermediate formed in the reaction catalyzed by *pyruvate dehydrogenase* C. Structure of intermediate formed in the reaction catalyzed by α-*ketoglutarate dehydrogenase*.

Thiamine pyrophosphate serves as a coenzyme in the formation or degradation of α-ketols by transketolase (Figure 12A), and in the oxidative decarboxylation of α-keto acids (Figure 12B).

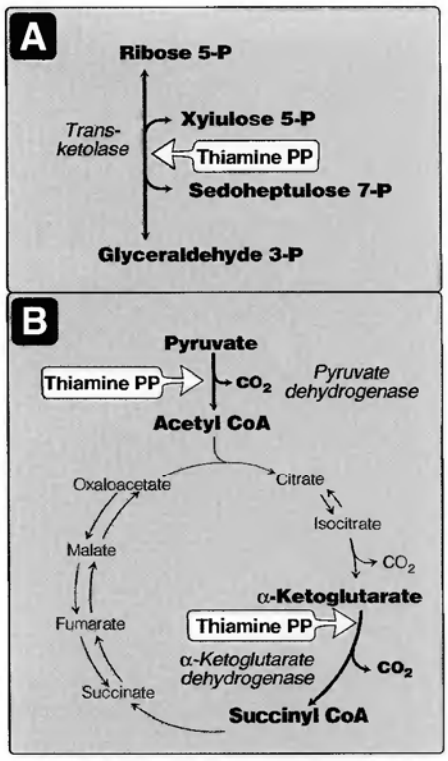


Figure 12
Reactions that use thiamine pyrophosphate (thiamine-PP) as coenzyme. A. *Transketolase*. B. *Pyruvate dehydrogenase* and *α-ketoglutarate dehydrogenase*.

Impaired cellular function:

1. Beriberi: This is a severe thiamine-deficiency syndrome found in areas where polished rice is the major component of the diet. Signs of infantile beriberi include tachycardia, vomiting, convulsions, and, if not treated, death. The deficiency syndrome can have a rapid onset in nursing infants whose mothers are deficient in thiamine. Adult beriberi is characterized by dry skin, irritability, disorderly thinking, and progressive paralysis.

2. Wernicke-Korsakoff syndrome: In the United States, thiamine deficiency, which is seen primarily in association with chronic alcoholism, is due to dietary insufficiency or impaired intestinal absorption of the vitamin. Some alcoholics develop Wernicke- Korsakoff syndrome--a deficiency state characterized by apathy, loss of memory, and a rhythmical to-and-fro motion of the eyeballs.

VI. NIACIN

Niacin, or nicotinic acid, is a substituted pyridine derivative. The biologically active coenzyme forms are nicotinamide adenine dinucleotide (NAD^+) and its phosphorylated derivative, nicotinamide adenine dinucleotide phosphate ($NADP^+$; Figure 13).

Nicotinamide, a derivative of nicotinic acid that contains an amide instead of a carboxyl group, also occurs in the diet. Nicotinamide is readily deaminated in the body and, therefore, is nutritionally equivalent to nicotinic acid. NAD^+ and $NADP^+$ serve as coenzymes in oxidation-reduction reactions in which the coenzyme undergoes reduction of the pyridine ring by accepting a hydride ion (hydrogen atom plus one electron; Figure 14). The reduced forms of NAD^+ and $NADP^+$ are NADH and NADPH, respectively.

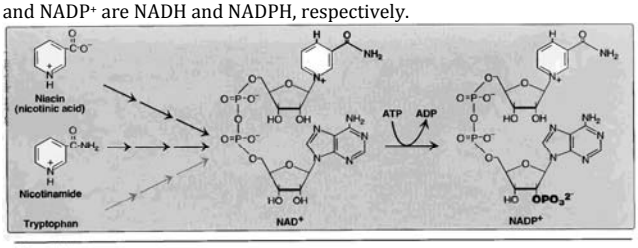


Figure 13
Structure and biosynthesis of NAD^+ and $NADP^+$.

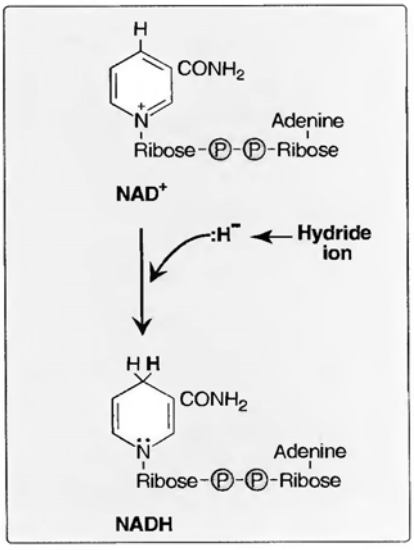


Figure 14 Reduction of NAD^+ to NADH.

A. Distribution of niacin

Niacin is found in unrefined and enriched grains and cereal, milk, and lean meats, especially liver. Limited quantities of niacin can also be obtained from the metabolism of tryptophan. [Note: The pathway is inefficient in that only about 1 mg of nicotinic acid is formed from 60 mg of tryptophan. Further, tryptophan is metabolized to niacin only when there is a relative abundance of the amino acid--that is, after
the needs for protein synthesis and energy production have been met.]

B. Clinical indications for niacin

1. Deficiency of niacin: A deficiency of niacin causes pellagra, a disease involving the skin, gastrointestinal (GI) tract, and CNS. The symptoms of pellagra progress through the three Ds: dermatitis, diarrhea, dementia, and, if untreated, death.
2. Treatment of hyperlipidemia: Niacin (at doses of 1.5 g/day or 100 times the RDA) strongly inhibits lipolysis in adipose tissue--the primary producer of circulating free fatty acids. The liver normally uses these circulating fatty acids as a major precursor for triacylglycerol synthesis. Thus, niacin causes a decrease in liver triacylglycerol synthesis, which is required for very-low-density lipoprotein (VLDL) production. Low-density lipoprotein (LDL, the cholesterol-rich lipoprotein) is derived from VLDL in the plasma. Thus, both plasma triacylglycerol (in VLDL) and cholesterol (in VLDL and LDL) are lowered. Therefore, niacin is particularly useful in the treatment of type lib hyperlipoproteinemia, which both VLDL and LDL are elevated.

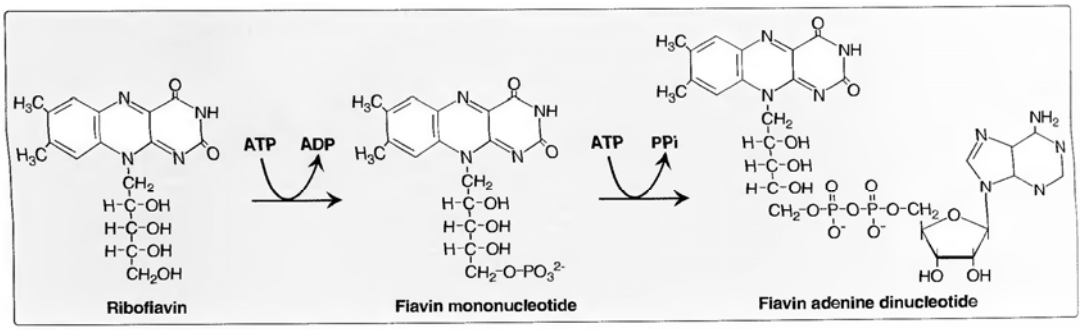


Figure 15 Structure and biosynthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD).

VII. RIBOFLAVIN (VITAMIN B2)

The two biologically active forms are flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), formed by the transfer of an AMP moiety from ATP to FMN (Figure 15).

FMN and FAD are each capable of reversibly accepting two hydrogen atoms, forming $FMNH_2$ or $FADH_2$. FMN and FAD are bound tightly--sometimes covalently--to flavoenzymes that catalyze the oxidation or reduction of a substrate.Riboflavin deficiency is not associated with a major human disease, although it frequently accompanies other vitamin deficiencies. Deficiency symptoms include dermatitis, cheilosis (fissuring at the corners of the mouth), and glossitis (the tongue appearing smooth and purplish).

VIII. BIOTIN

Biotin is a coenzyme in carboxylation reactions, in which it serves as a carrier of activated carbon dioxide.

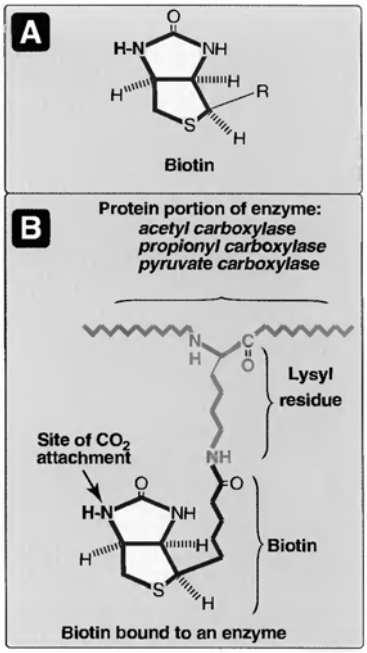


Figure 16
A. Structure of biotin. B. Biotin covalently bound to a lysyl residue of a biotin-dependent enzyme.

Biotin is covalently bound to the ε-amino groups of lysine residues of biotin-dependent enzymes (Figure 16). Biotin deficiency does not occur naturally because the vitamin is widely distributed in food. Also, a large percentage of the biotin requirement in humans is supplied by intestinal bacteria. However, the addition of raw egg-white to the diet as a source of protein induces symptoms of biotin deficiency, namely, dermatitis, glossitis, loss of appetite, and nausea. Raw egg white contains a glycoprotein, avidin, which tightly binds biotin and prevents its absorption from the intestine. However, with a normal diet, it has been estimated that 20 eggs per day would be required to induce a deficiency syndrome. Thus, inclusion of an occasional raw egg in the diet does not lead to biotin deficiency.

IX. PANTOTHENIC ACID

Pantothenic acid is a component of coenzyme A, which functions in the transfer of acyl groups (Figure 17). Coenzyme A contains a thiol group that carries acyl compounds as activated thiol esters. Examples of such structures are succinyl CoA, fatty acyl CoA, and acetyl CoA. Pantothenic acid is also a component of fatty acid synthase. Eggs, liver, and yeast are the most important sources of pantothenic acid, although the vitamin is widely distributed. Pantothenic acid deficiency is not well characterized in humans, and no RDA has been established.

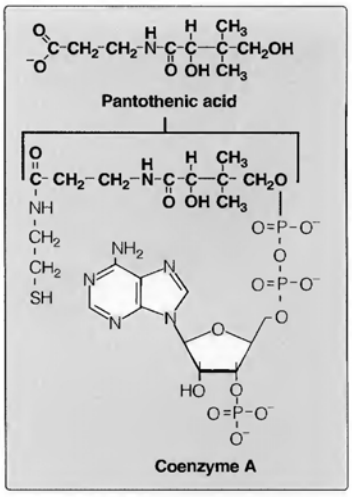


Figure 17 Structure of coenzyme A.

A. Structure of vitamin A

Vitamin A is often used as a collective term for several related biologically active molecules (Figure 18). The term retinoids includes both natural and synthetic forms of vitamin A that may or may not show vitamin A activity.

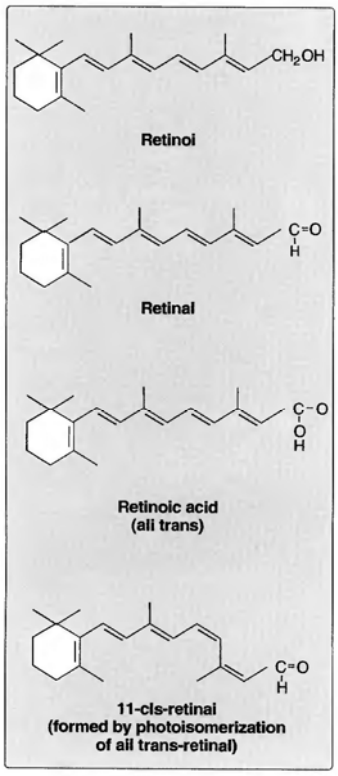


Figure 18 Structure of the retinoids.

1. Retinol: A primary alcohol containing a β-ionone ring with an unsaturated side chain, retinol is found in animal tissues as a retinyl ester with long-chain fatty acids.
2. Retinal: This is the aldehyde derived from the oxidation of retinol. Retinal and retinol can readily be interconverted.
3. Retinoic acid: This is the acid derived from the oxidation of retinal. Retinoic acid cannot be reduced in the bod, and, therefore, cannot give rise to either retinal or retinol.
4. β-carotene: Plant foods contain β-carotene, which can be oxidatively cleaved in the intestine to yield two molecules of retinal. In humans, the conversion is inefficient, and the vitamin A activity of β-carotene is only about one sixth that of retinol.

B. Absorption and transport of vitamin A

1. Transport to the liver: Retinol esters present in the diet are hydrolyzed in the intestinal mucosa, releasing retinol and free fatty acids (Figure 19). Retinol derived from esters and from the cleavage and reduction of carotenes is reesterified to long-chain fatty acids in the intestinal mucosa and secreted as a component of chylomicrons into the lymphatic system (see Figure 19). Retinol esters contained in chylomicrons are taken up by, and stored in, the liver.

2. Release from the liver: When needed, retinol is released from the liver and transported to extrahepatic tissues by the plasma retinol-binding protein (RBP). The retinol-RBP complex attaches to specific receptors on the surface of the cells of peripheral tissues, permitting retinol to enter. Many tissues contain a cellular retinol-binding protein that carries retinol to sites in the nucleus where the vitamin acts in a manner analogous to steroid hormones.

C. Mechanism of action of vitamin A

Retinoic acid binds with high-affinity to specific receptor proteins present in the nucleus of target tissues, such as epithelial cells (Figure 20). The activated retinoic acid-receptor complex interacts with nuclear chromatin to stimulate retinoid-specific RNA synthesis, resulting in the production of specific proteins that mediate several physiologic functions. For example, retinoids control the expression of the keratin gene in most epithelial tissues of the body.

The specific retinoic acid-receptor proteins are part of the superfamily of transcriptional regulators that includes the steroid and thyroid hormones and 1,25-dihydroxycholecalciferol, all of which function in a similar way.

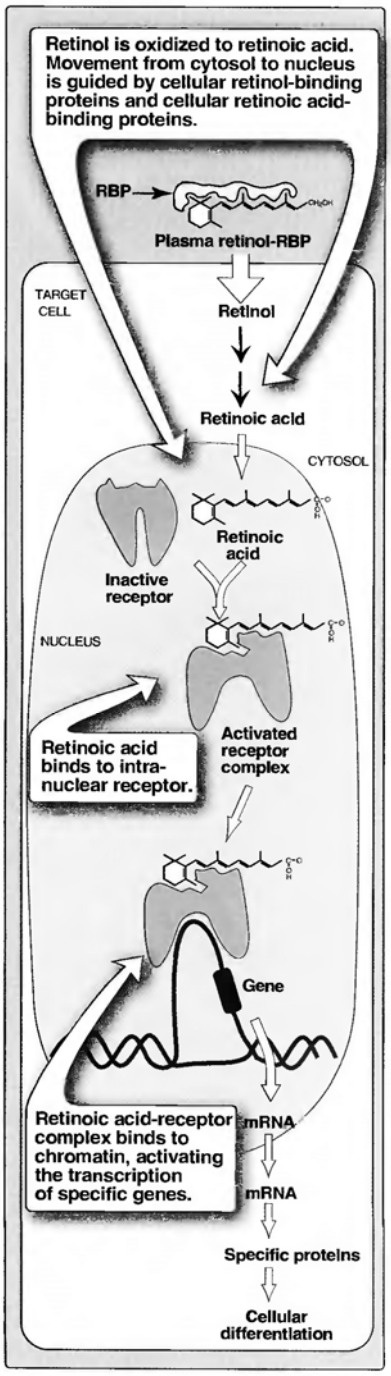


Figure 20 Action of retinoids

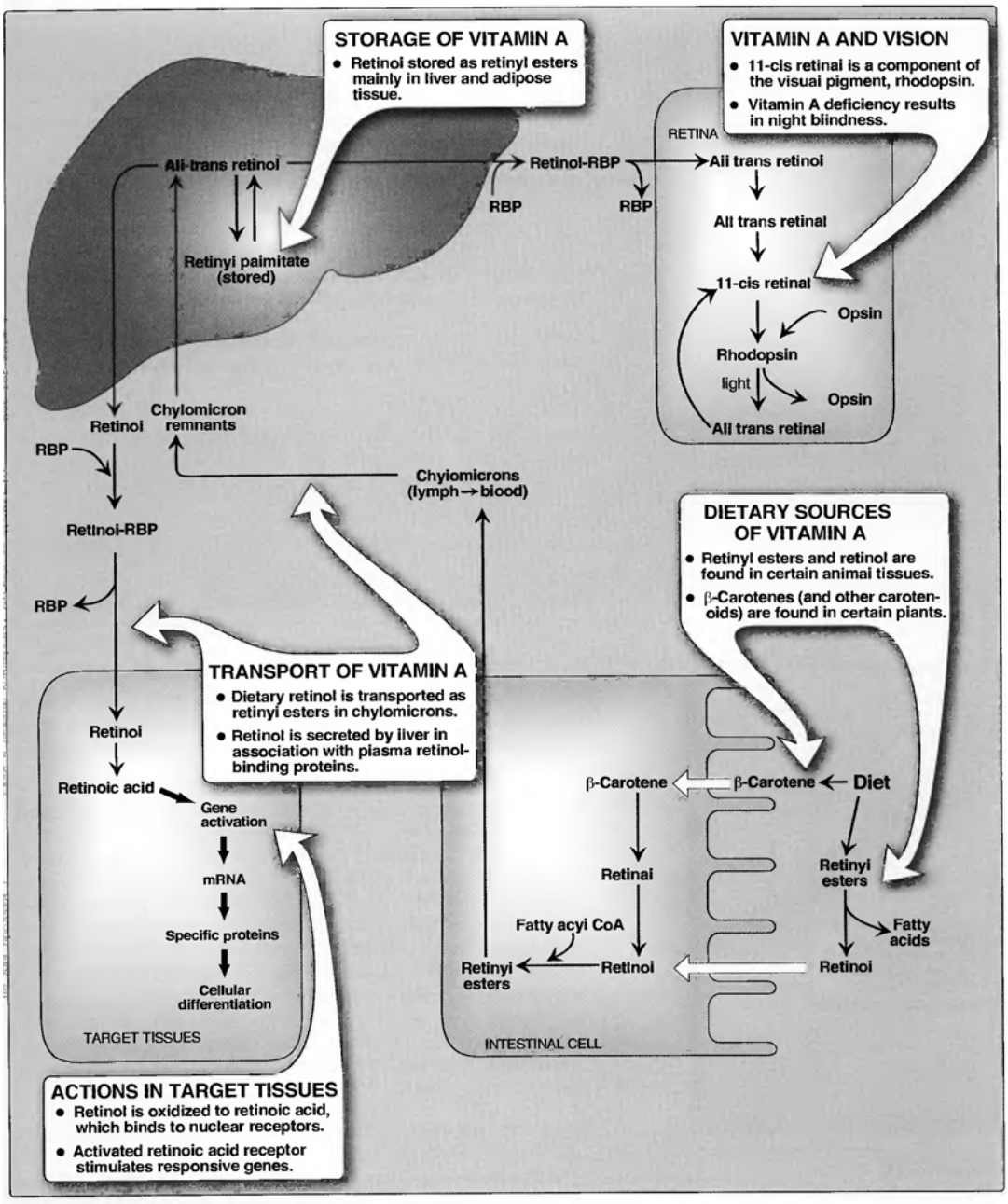


Figure 19
Absorption, transport, and storage of vitamin A and its derivatives. RBP = retinol-binding protein.

D. Functions of vitamin A

1. Visual cycle: Vitamin A is a component of the visual pigments of rod and cone cells. Rhodopsin, the visual pigment of the rod cells in the retina, consists of 11-cis retinal specifically bound to the protein opsin. When rhodopsin is exposed to light, a series of photochemical isomerizations occurs, which results in the bleaching of the visual pigment and release of all trans retinal and opsin.

This process triggers a nerve impulse that is transmitted by the optic nerve to the brain. Regeneration of rhodopsin requires isomerization of all trans retinal back to 11-cis retinal. Trans retinal, after being released from rhodopsin, is isomerized to 11-cis retinal, which spontaneously combines with opsin to form rhodopsin, thus completing the cycle. Similar reactions are responsible for color vision in the cone cells.

2. Growth: Animals deprived of vitamin A initially lose their appetites, possibly because of keratinization of the taste buds. Bone growth is slow and fails to keep pace with growth of the nervous system, leading to central nervous system damage.

3. Reproduction: Retinol and retinal are essential for normal reproduction, supporting spermatogenesis in the male and preventing fetal resorption in the female. Retinoic acid is inactive in maintaining reproduction and in the visual cycle, but promotes growth and differentiation of epithelial cells; thus, animals given vitamin A only as retinoic acid from birth are blind and sterile.

4. Maintenance of epithelial cells: Vitamin A is essential for normal differentiation of epithelial tissues and mucus secretion.

D. Distribution of vitamin A:
Liver, kidney, cream, butter, and egg yolk are good sources of preformed vitamin A. Yellow and dark green vegetables and fruits are good dietary sources of the carotenes, which serve as precursors of vitamin A.

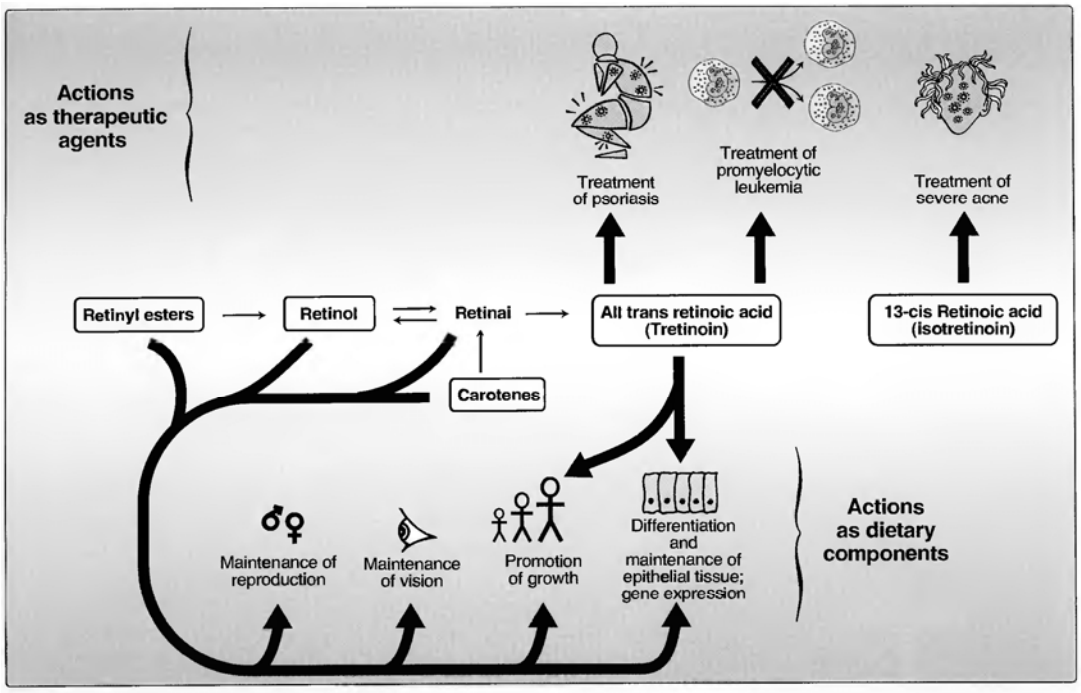


Figure 21
Summary of actions of retinoids. Compounds in boxes are available as dietary components or as pharmacologic agents.

E. Requirement for vitamin A

The RDA for adults is 1000 retinol equivalents (RE) for males and 800 RE for females. One RE = 1 mg of retinol, 6 mg of β-carotene, or 12 mg of other carotenoids.

F. Clinical indications

Although chemically related, retinoic acid and retinol have distinctly different therapeutic applications. Retinol and its precursor are used as dietary supplements, whereas various forms of retinoic acid are
useful in dermatology.

1. Dietary deficiency: Vitamin A, administered as retinol or retinyl esters, is used to treat patients deficient in the vitamin (Figure 21). Night blindness is one of the earliest signs of vitamin A deficiency. The visual threshold is increased, making it difficult to see in dim light. Prolonged deficiency leads to an irreversible loss in the number of visual cells. Severe vitamin A deficiency leads to xerophthalmia, a pathologic dryness of the conjunctiva and cornea. If untreated, xerophthalmia results in corneal ulceration and, ultimately, in blindness because of the formation of opaque scar tissue. The condition is most frequently seen in children in developing tropical countries. Over 500,000 children worldwide are blinded each year by xerophthalmia caused by insufficient vitamin A in the diet.

2. Acne and psoriasis: Dermatologic problems such as acne and psoriasis are effectively treated with retinoic acid or its derivatives (see Figure 21). Mild cases of acne, Darier disease, and skin aging are treated with topical application of tretinoin (all trans retinoic acid), as well as benzoyl peroxide and antibiotics. [Note: Tretinoin is too toxic for systemic administration and is confined to topical application.] In patients with severe recalcitrant cystic ache unresponsive to conventional therapies, the drug of choice is isotretinoin (13-cis retinoic acid) administered orally.

3. Prevention of chronic disease: Populations consuming diets high in β-carotene show decreased incidence of heart disease and lung and skin cancer (see Figure 21). Consumption of foods rich in β-carotene is also associated with reduced risk of cataracts and macular degeneration. However, in clinical trials, β-carotene supplementation not only did not decrease the incidence of lung cancer, but actually increased cancer in individuals who smoke. Subjects in a clinical trial who received high doses of β-carotene unexpectedly had increased death due to heart disease.

6. Toxicity of retinoids

1. Vitamin A: Excessive intake of vitamin A produces a toxic syndrome called hypervitaminosis A. Amounts exceeding 7.5 mg/day of retinol should be avoided. Early signs of chronic hyper vitaminosis A are reflected in the skin, which becomes dry and pruritic, the liver, which becomes enlarged and can become cirrhotic, and in the nervous system, where a rise in intracranial pressure may mimic the symptoms of a brain tumor. Pregnant women particularly should not ingest excessive quantities of vitamin A because of its potential for causing congenital malformations in the developing fetus.

2. Isotretinoin: The drug is teratogenic and absolutely contraindicated in women with childbearing potential unless they have severe, disfiguring cystic acne that is unresponsive to standard therapies. Pregnancy must be

excluded before initiation of treatment, and adequate birth control must be used. Prolonged treatment with isotretinoin leads to hyperlipidemia and an increase in the LDL/HDL ratio, providing some concern for an increased risk of cardiovascular disease.

XI. VITAMIN D

The D vitamins are a group of sterols that have a hormone-like function. The active molecule, 1,25-dihydroxycholecalciferol (1,25 diOH D3), binds to intracellular receptor proteins. The 1,25-diOH D3-receptor complex interacts with DNA in the nucleus of target cells in a manner similar to that of vitamin A (see Figure 20), and either selectively stimulates gene expression, or specifically represses gene transcription. The most prominent actions of 1,25-diOH D3 are to regulate the plasma levels of calcium and phosphorus.

A. Distribution of vitamin D

1. Diet: Ergocalciferol (vitamin D2), found in plants, and cholecalciferol (vitamin D3), found in animal tissues, are sources of preformed vitamin D activity (Figure 22). Ergocalciferol and cholecalciferol differ chemically only in the presence of an additional double bond and methyl group in the plant sterol.

2. Endogenous vitamin precursor: 7-Dehydrocholesterol, an intermediate in cholesterol synthesis, is converted to cholecalciferol in the dermis and epidermis of humans exposed to sunlight. Preformed vitamin D is a dietary requirement only in individuals with limited exposure to sunlight.

B. Metabolism of vitamin D

1. Formation of 1,25-diOH D3: Vitamins D2 and D3 are not biologically active, but are converted in vivo to the active form of the D vitamin by two sequential hydroxylation reactions (Figure 23). The first hydroxylation occurs at the 25-position, and is catalyzed by a specific hydroxylase in the liver. The product of the reaction, 25-hydroxycholecalciferol (25-OH D3), is the predominant form of vitamin D in the plasma and the major storage form of the vitamin. 25-OH D3 is further hydroxylated at the one position by a specific 25-hydroxycholecalciferol 1-hydroxylase found primarily in the kidney, resulting in the formation of 1,25-dihydroxycholecalciferol (1,25-diOH D3). [Note: This hydroxylase, as well as the liver 25-hydroxylase, employ cytochrome P450, molecular oxygen, and NADPH.]

2. Regulation of 25-hydroxycholecalciferol 1-hydroxylase: 1,25-diOH D3 is the most potent vitamin D metabolite. Its formation is tightly regulated by the level of plasma phosphate and calcium ions (Figure 24). 25-Hydroxycholecalciferol 1-hydroxylase activity is increased directly by low plasma phosphate or indirectly by low plasma calcium, which triggers the release of parathyroid hormone (PTH). Hypocalcemia caused by insufficient dietary calcium thus results in elevated levels of plasma 1,25 diOH D3. 1-Hydroxylase activity is also decreased by excess 1,25 diOH D3, the product of the reaction.

C. Function of vitamin D

The overall function of 1,25-diOH D3 is to maintain adequate plasma levels of calcium. It performs this function by: 1) increasing uptake of calcium by the intestine, 2) minimizing loss of calcium by the kidney, and 3) stimulating resorption of bone when necessary (see Figure 23).

1. Effect of vitamin D on the intestine: 1,25-diOH D3 stimulates intestinal absorption of calcium and phosphate. 1,25-diOH D3 enters the intestinal cell and binds to a cytosolic receptor. The 1,25-diOH D3-receptor complex then moves to the nucleus where it selectively interacts with the cellular DNA. As a result, calcium uptake is enhanced by an increased synthesis of a specific calcium-binding protein. Thus, the mechanism of action of 1,25-diOH D3 is typical of steroid hormones.

2. Effect of vitamin D on bone: 1,25-diOH D3 stimulates the mobilization of calcium and phosphate from bone by a process that requires protein synthesis and the presence of PTH. The result is an increase in plasma calcium and phosphate. Thus, bone is an important reservoir of calcium that can be mobilized to maintain plasma levels.

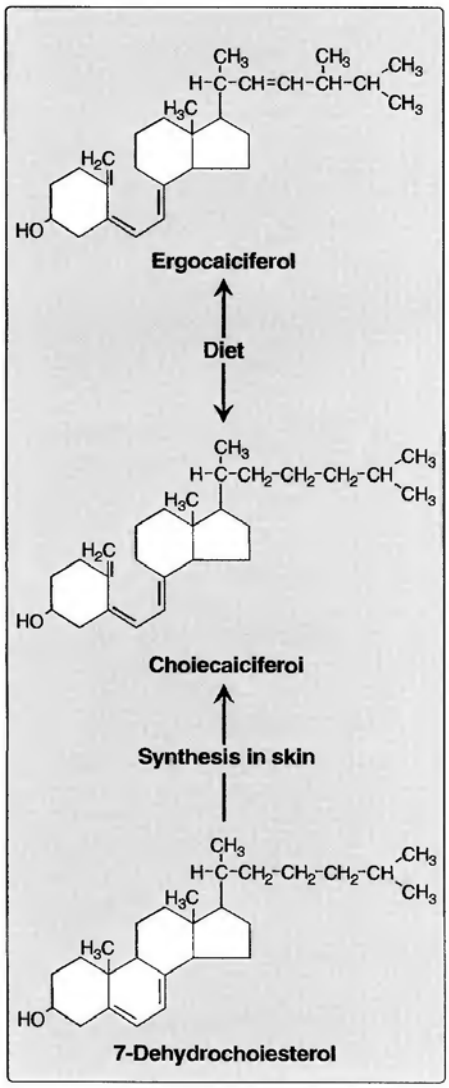


Figure 22 Sources of vitamin D.

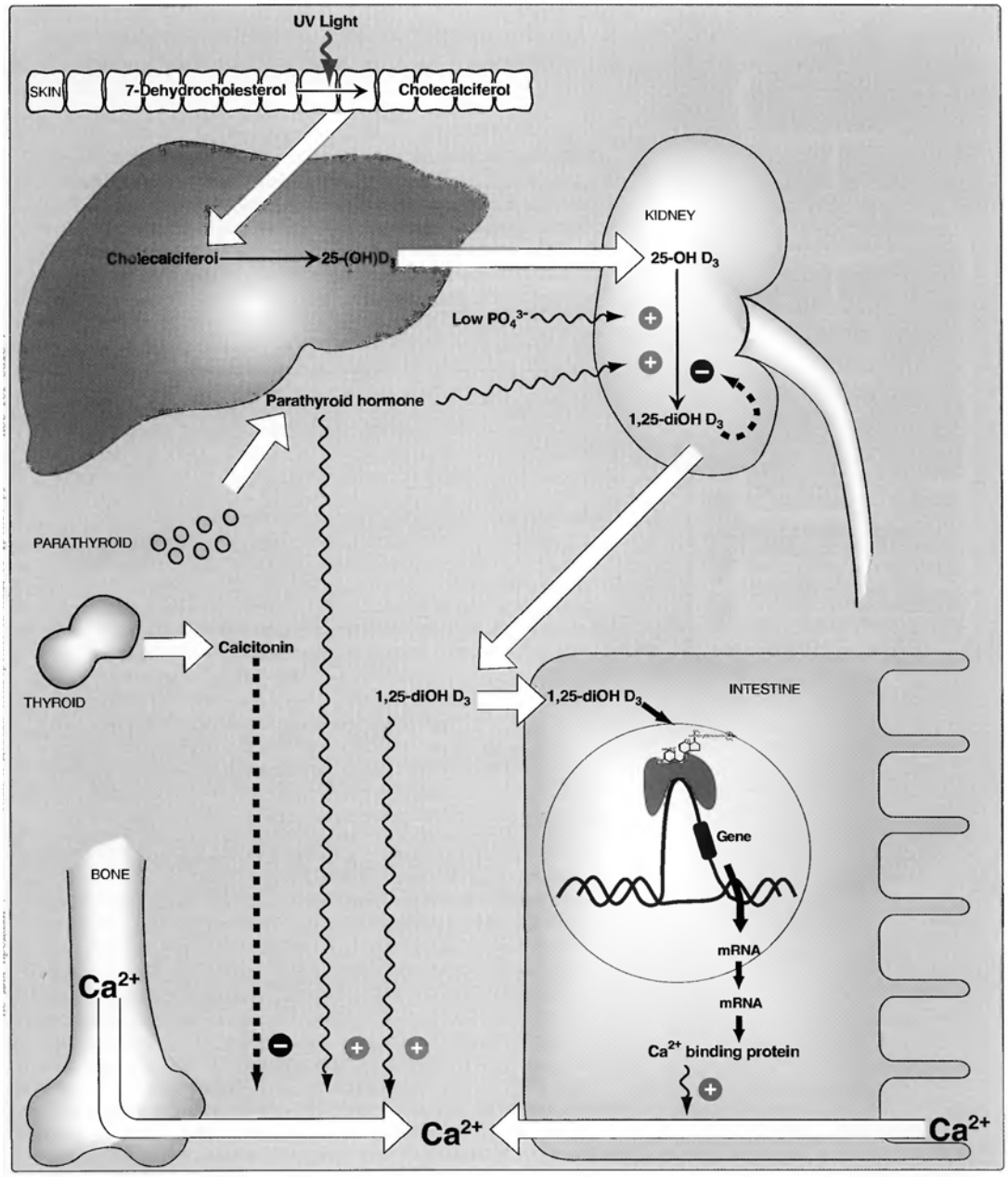


Figure 23 Metabolism and actions of vitamin D.

D. Distribution and requirement of vitamin D

Vitamin D occurs naturally in fatty fish, liver, and egg yolk. Milk, unless it is artificially fortified, is not a good source of the vitamin. The RDA for adults is 5 mg cholecalciferol, or 200 international units (IU) of vitamin D.

E. Clinical indications

1. Nutritiona| rickets: Vitamin D deficiency causes a net demineralization of bone, resulting in rickets in children and osteomalacia in adults (Figure 25). Rickets is characterized by the continued formation of the collagen matrix of bone, but incomplete mineralization, resulting in soft, pliable bones. In osteomalacia, demineralization of preexisting bones increases their susceptibility to fracture. Insufficient exposure to daylight and/or deficiencies in vitamin D consumption occur predominantly in infants and the elderly. Vitamin D deficiency is more common in the northern latitudes, because less vitamin D synthesis occurs in the skin as a result of reduced exposure to ultraviolet light. [Note: The RDA of 200 IU/day (which corresponds to 5 ,ug of chole- calciferol) may be insufficient, because higher doses of 800 IU/day have been shown to reduce the incidence of osteoporotic fractures.]

2. Renal rickets (renal osteodystrophy): This disorder results from chronic renal failure and, thus, the decreased ability to form the active form of the vitamin. 1,25-diOH cholecalciferol (calcitriol) administration is effective replacement therapy.

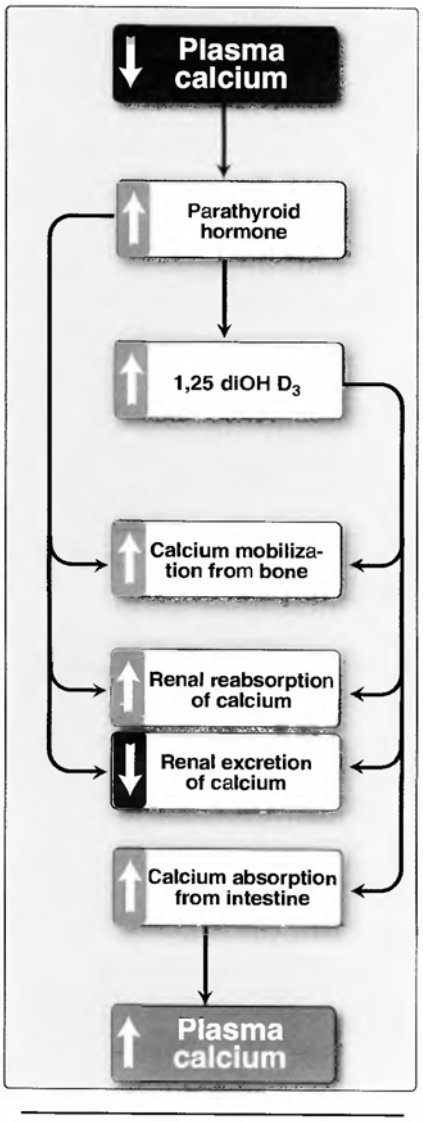


Figure 24 Response to low plasma calcium.

3. Hypoparathyroidism: Lack of parathyroid hormone causes hypocalcemia and hyperphosphatemia. These patients may be treated with any form of vitamin D, together with parathyroid hormone.

F. Toxicity of vitamin D

Vitamin D is the most toxic of all vitamins. Like all fat-soluble vitamins, vitamin D can be stored in the body and is only slowly metabolized. High doses (100,000 IU for weeks or months) can cause loss of appetite, nausea, thirst, and stupor. Enhanced calcium absorption and bone resorption results in hypercalcemia, which can lead to deposition of calcium in many organs, particularly the arteries and kidneys.

XII. VITAMIN K

The principal role of vitamin K is in the post-translational modification of various blood clotting factors, in which it serves as a coenzyme in the carboxylation of certain glutamic acid residues present in these pro.t.eins. Vitamin K exists in several forms, for example, in plants as phylloquinone (or vitamin K1), and in intestinal bacterial flora as menaquinone (or vitamin K2). For therapy, a synthetic derivative of vitamin K, menadione, is available.

A. Function of vitamin K

1. Formation of y-carboxyglutamate: Vitamin K is required in the hepatic synthesis of prothrombin and blood clotting factors II, VII, IX, and X. These proteins are synthesized as inactive precursor molecules. Formation of the clotting factors requires the vitamin K-dependent carboxylation of glutamic acid residues (Figure 26). This forms a mature clotting factor that contains -y-carboxyglutamate (Gla) and is capable of subsequent activation. The reaction requires O_2, CO_2, and the hydroquinone form of vitamin K. The formation of Gla is sensitive to inhibition by dicumaroi, an anticoagulant occurring naturally in spoiled sweet clover, and by warfarin, a synthetic analog of vitamin K.

2. Interaction of prothrombin with platelets: The Gla residues of prothrombin are good chelators of positively charged calcium ions, because of the two adjacent, negatively charged carboxylate groups.. The prothrombin-calcium complex is then able to bind to phospholipids essential for blood clotting on the surface of platelets. Attachment to the platelet increases the rate at which the proteolytic conversion of prothrombin to thrombin can occur (Figure 27).

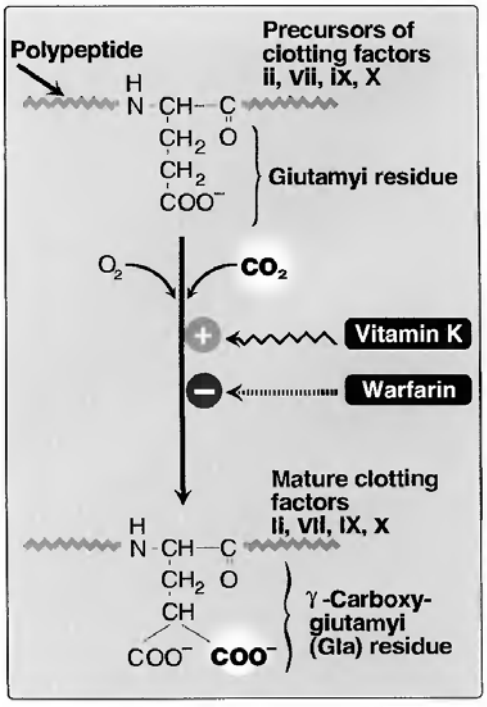


Figure 26
Carboxylation of glutamate to form γ-carboxyglutamate (Gla).

3. Role of y-carboxyglutamate residues in other proteins: Gla is also present in other proteins (for example, osteocalcin of bone) unrelated to the clotting process. However, the physiologic role of these proteins and the function of vitamin K in their synthesis is not yet understood.

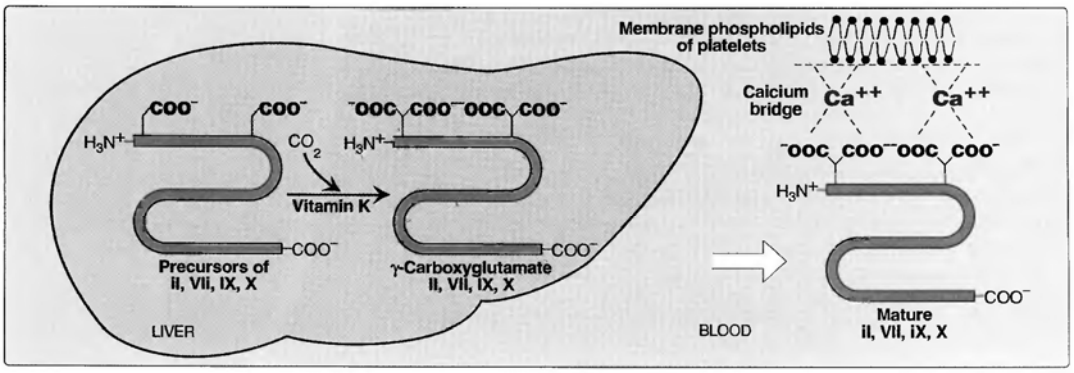


Figure 27 Role of vitamin K in blood coagulation.

B. Distribution and requirement of vitamin K

Vitamin K is found in cabbage, cauliflower, spinach, egg yolk, and liver. There is also extensive synthesis of the vitamin by the bacteria in the gut. There is no RDA for vitamin K, but 70 to 140 rag/day is recommended as an adequate level. The lower level assumes one half of the estimated requirement comes from bacterial synthesis whereas the upper figure assumes no bacterial synthesis.

C. Clinical indications

1. Deficiency of vitamin K: A true vitamin K deficiency is unusual because adequate amounts are generally produced by intestinal bacteria or obtained from the diet. If the bacterial population in the gut is decreased, for example by antibiotics, the amount of endogenously formed vitamin is depressed, and can lead to hypoprothrombinemia in the marginally malnourished individual (for example, a debilitated geriatric patient). This condition may require supplementation with vitamin K to correct the bleeding tendency. In addition, certain second generation cephalosporins (for example, cefoperazone, cefamandole, and moxalactam) cause hypoprothrombinemia, apparently by a warfarin-like mechanism. Consequently, their use in treatment is usually supplemented with vitamin K.

2. Deficiency of vitamin K in the newborn: Newborns have sterile intestines and cannot initially synthesize vitamin K. Because human milk provides only about one fifth of the daily requirement for vitamin K, it is recommended that all newborns receive a single intramuscular dose of vitamin K as prophylaxis against hemorrhagic disease.

D Toxicity of vitamin K

Prolonged administration of large doses of vitamin K can producehemolytic anemia and jaundice in the infant, due to toxic effects on the membrane of red blood cells.

XIII. VITAMIN E

The E vitamins consist of eight naturally occurring tocopherols, of which (y-tocopherol is the most active (Figure 28). The primary function of vitamin E is as an antioxidant in prevention of the nonenzymic oxidation of cell components (for example, polyunsaturated fatty acids) by molecular oxygen and free radicals.

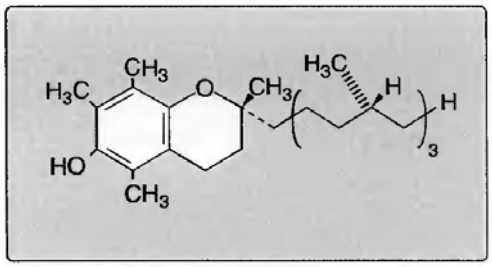


Figure 28 Structure of vitamin E.

A. Distribution and requirements of vitamin E

Vegetable oils are rich sources of vitamin E, whereas liver and eggs contain moderate amounts. The RDA for tocopherol is 10 mg for men and 8 mg for women. Vitamin E requirement increases as the intake of polyunsaturated fatty acid increases.

B. Deficiency of vitamin E

Vitamin E deficiency is almost entirely restricted to premature to infants. When observed in adults, it is usually associated with defective lipid absorption or transport. The signs of human vitamin E deficiency include sensitivity of erythrocytes to peroxide, and the appearance of abnormal cellular membranes.

C. Clinical indications

Vitamin E is not recommended for the prevention of chronic disease, such as coronary heart disease or cancer. Clinical trials using vitamin E supplementation have been uniformly disappointing. For example, subjects in the Alpha-Tocopherol, Beta Carotene Cancer Prevention Study trial who received high doses of vitamin E, not only lacked cadiovascular benefit but also had an increased incidence of stroke

D. Toxicity of vitamin E

Vitamin E is the least toxic of the fat-soluble vitamins, and no toxicity has been observed at doses of 300 mg/day.

Figure 29

VITAMIN	OTHER NAMES	ACTIVE FORM	FUNCTION
Folic acid	—	Tetrahydro-folic acid	Transfer one-carbon units; Synthesis of methionine, purines, and thymine
Vitamin B_{12}	Cobalamin	Methylcobalamin Deoxyadenosyl cobalamin	Cofactor for reactions: Homocysteine → Methionine Methylmalonyl CoA → Succinyl CoA
Vitamin C	Ascorbic acid	Ascorbic acid	Antioxidant Cofactor for hydroxylation reactions, for example: In procollagen: Proline → hydroxyproline Lysine → hydroxylysine
Vitamin B_6	Pridoxine Pyridoxamine Pyridoxal	Pyridoxal phosphate	Cofactor for enzymes, particularly in amino acid metabolism
Vitamin B_1	Thiamine	Thiamine pyrophosphate	Cofactor of enzymes catalyzing: Pyruvate → acetyl CoA α-Ketoglutarate → Succinyl CoA Ribose 5-P xylulose 5-P → Sedoheptulose 7-P + Glyceraldehyde 3-P
Niacin	Nicotinic acid Nicotinamide	NAD^+, $NADP^+$	Electron transfer
Vitamin B_2	Riboflavin	FMN, FAD	Electron transfer
Biotin	—	Enzyme-bound biotin	Carboxylation reactions
Pantothenic acid	—	Coenzyme A	Acyl carrier
			WATER-SOLUBLE
			FAT-SOLUBLE
Vitamin A	Retinol Retinal Retinoic acid β-Carotene	Retinol Retinal Retinoic acid	Maintenance of reproduction Vision Promotion of growth Differentiation and maintenance of epithelial tissues Gene expression
Vitamin D	Cholecalciferol Ergocalciferol	1,25-Dihydroxy-cholecalciferol	Calcium uptake
Vitamin K	Menadione Menaquinone Phylloquinone	Menadione Menaquinone Phylloquinon	γ-Carboxylation of glutamate residue in clotting and other proteins
Vitamin E	α-Tocopherol	Any of several tocopherol derivatives	Antioxidant

DEFICIENCY	SIGNS AND SYMPTOMS	TOXICITY	NOTES
Megaloblastic anemia Neural tube defects	Anemia Birth defects	None	Administration of high levels of folate can mask vitamin B_{12} deficiency.
Pernicious anemia Dementia Spinal degeneration	Megaloblastic anemia Neuropsychiatric symptoms	None	Parnicious anemia is treated with IM or high-dose oral vitamin B_{12}
Scurvy	Sore, spongy gums Loose teeth Poor wound healing	None	Benefits of supplementation not established in controlled trials
Rare	Glossitis Neuropathy	Yes	Deficiency can be induced by isoniazid Sansory neuropathy occurs at high doses
Beriberi Wernicke-Korsakoff syndrome (most common in alcoholics)	Tachycardia, vomiting, convulsions Apathy, loss of memory, eye movements	None	—
Pellagra	Dermatitis Diarrhea Dementia	None	High doses of niacin used to treat hyperlipidemia
Rare	Dermatitis Angular stomatitis	None	—
Rare	—	None	Consumption of large amounts of raw egg whites (which contains a protein, avidin, that binds biotin) can induce a biotin deficiency
Rare	—	None	— **WATER-SOLUBLE**
			FAT-SOLUBLE
Impotence Night blindness Retardation of growth Xerophthalmia	Increased visual threshold Dryness of cornea	Yes	β-Carotene not acutely toxic, but supplementation is not recommended Excess vitamin A can increase incidence of fractures
Rickets (in children) Osteomalacia (in adults)	Soft, pliable bones	Yes	Vitamin D is not a true vitamin because it can be synthesized in skin. Application of sunscreen lotions or presence of dark skin color decreases this synthesis.
Newborn Rare in adults	Bleeding	Rare	Vitamin K produced by intestinal bacteria. Vitamin K deficiency common in newborns Parenteral treatment with the vitamin K is recommended at birth
Rare	Red blood cell fragility leads to hemolytic enemia	None	Benefits of supplementation not established in controlled trials

UNIT-2

CELLULAR ORGANIZATION

INSTITUTE FOR ADVANCED STUDIES

B-7 SARASWATI NAGAR, JODHPUR

e-mail:ifasnet@gmail.com

http://csirnetlifesciences.tripod.com

BLANK PAGE

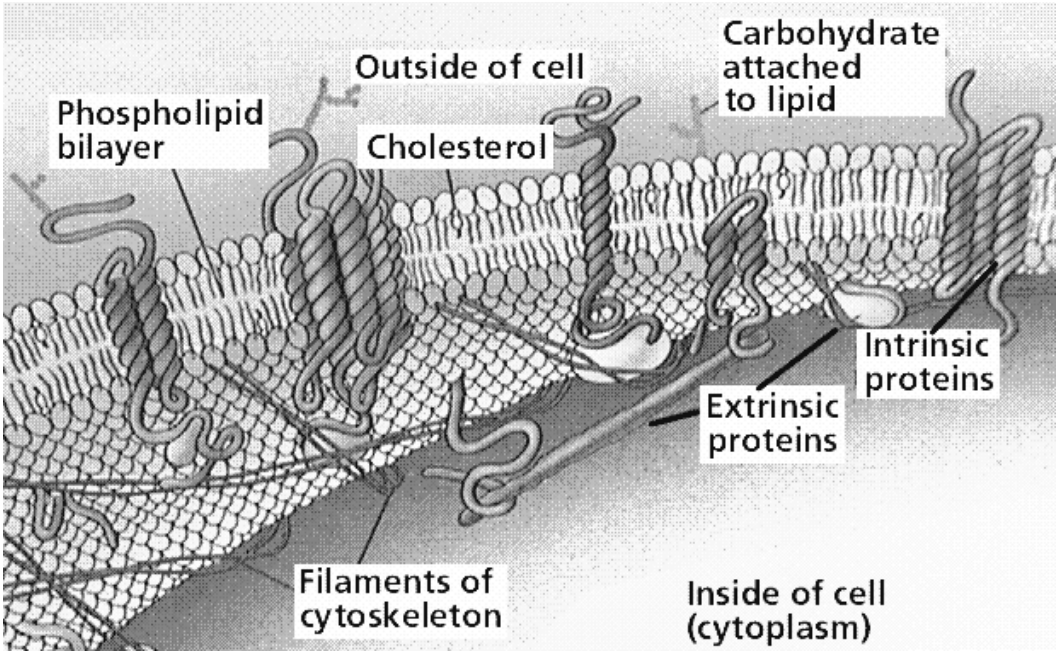
A1. MEMBRANES-STUCTURE

Cell membranes are crucial to the life of the cell. The plasma membrane encloses the cell, defines its boundaries, and maintains the essential differences between the cytosol and the extracellular environment. Inside the cell the membranes of the endoplasmic reticulum, Golgi apparatus, mitochondria, and other membrane-bounded organelles in eucaryotic cells maintain the characteristic differences between the contents of each organelle and the cytosol. Ion gradients across membranes, established by the activities of specialized membrane proteins, can be used to synthesize ATP, to drive the transmembrane movement of selected solutes, or, in nerve and muscle cells, to produce and transmit electrical signals. In all cells the plasma membrane also contains proteins that act as sensors of external signals, allowing the cell to change its behavior in response to environmental cues; these protein sensors, or receptors, transfer information rather than ions or molecules across the membrane.

MEMBRANE ULTRASTRUCTURE AND FLUID MOSAIC MODEL

This model, most forcefully proposed by S.J. Singer (A.R. Biochem. 1974, 43: 805-826), is the currently most favored model. Some consider it the 6 most current modification of the original Danielli-Davson model. It proposes:
1. A phospholipid bilayer with the hydrophobic (nonpolar) fatty acid chains directed inwards and the polar hydrophilic heads lining the surfaces.

2. Proteins, glycoproteins, and lipoproteins, associate with the inner and outer surfaces of the membranes. The polar portions of the proteins associate with the polar surface and may protrude from it while portions of the proteins may be embedded in the bilayer of lipid. It is proposed that the nonpolar portions of these proteins associate with the nonpolar fatty acid chains. Proteins having both polar and nonpolar areas (residues) are termed amphipathic.



Despite their differing functions, all biological membranes have a common general structure: each is a very thin film of lipid and protein molecules, held together mainly by non-covalent interactions. Cell membranes are dynamic, fluid structures, and most of their molecules are able to move about in the plane of the membrane. The lipid molecules are arranged as a continuous double layer about 5 nm thick. This lipid bilayer provides the basic structure of the membrane and serves as a relatively impermeable barrier to the passage of most water-soluble molecules. Protein molecules "dissolved" in the lipid bilayer mediate most of the other functions of the membrane, transporting specific molecules across it, for example, or catalyzing membrane-associated reactions, such as ATP synthesis. In the plasma membrane some proteins serve as structural links that connect the membrane to the cytoskeleton and/or to either the extracellular matrix or an adjacent cell, while others serve as receptors to detect and transduce chemical signals in the cell's environment. As would be expected, cell membranes are asymmetrical structures: the lipid and protein compositions of the outside and inside faces differ from one another in ways that reflect the different functions performed at the two surfaces of the membrane.

3. Under physiological conditions the membrane components are not rigidly fixed, but exist in a fluid and dynamic state. Lipids of the membrane are fluid at physiological temperatures, thus there may be fluid movement of the lipid molecules and lateral movement (diffusion) of the proteins that are embedded in the membrane. "Protein icebergs in a sea of lipid." (Singer) This forms a heterogeneous mosaic of proteins which may be continuously rearranged (but often with specificity.)

4. Not only is there lateral mobility of lipids but there is also exchange from one monolayer of the bi-leaflet to the other monolayer. This exchange (flip-flop) is much slower than the lateral exchange, detected by nuclear resonance studies. This process is facilitated by specific proteins (enzymes) called flippases.

5. Proteins may be integral or peripheral, probably most being integral and displaying various different secondary and tertiary forms. The proteins, being capable of secondary, tertiary, and quaternary structure changes in response to stimuli, may reorient in relation to the phospholipids bilayer.

6. In a phospholipid bilayer, the long fatty acyl side chains in each leaflet are oriented toward one another, forming a hydrophobic core; the polar head groups line both surfaces.

7. The phospholipid bilayer forms the basic structure of all biomembranes, which also contain proteins, glycoproteins, cholesterol and other steroids, and glycolipids. The presence of specific sets of membrane proteins permits each type of membrane to carry out distinctive functions.

8. All cellular membranes line closed compartments and have a cytosolic and an exoplasmic face. The asymmetry of biological membranes is reflected in the specific orientation of each type of integral and peripheral membrane protein with respect to the cytosolic and exoplasmic faces. The presence of glycolipids exclusively in the exoplasmic leaflet also contributes to membrane asymmetry.

9. Most integral proteins and lipids are laterally mobile in biomembranes. According to the fluid mosaic model, the membrane is viewed as a two-dimensional mosaic of phospholipid and protein molecules.

10. As a phospholipid bilayer is heated, it undergoes a phase transition from a gel-like to a more fluid state over a short temperature range.

11. Cholesterol is a major determinant of bilayer fluidity, although its effect depends on the composition of a membrane. Natural biomembranes generally have a fluid-like consistency, and cells adjust their phospholipid composition to maintain bilayer fluidity.

12. In all cells, proteins in the plasma membrane selectively absorb nutrients, expel wastes, and maintain the proper intracellular ionic composition. Proteins in the plasma membrane anchor the membrane to intracellular cytoskeletal fibers and the extracellular matrix or cell wall. In multicellular organisms, plasma membrane proteins also act in the interactions and communication between cells, which are critical for proper functioning of multicellular tissues.

13. In plants, the cell wall, which is built mainly of cellulose, is the major determinant of cell shape and imparts rigidity to cells. Animal cells, which lack a wall, are surrounded by an extracellular matrix consisting of collagen, glycoproteins

A2. MEMBRANE PROTEINS

1. Transmembrane Proteins:

Membrane proteins are either extrinsic or intrinsic. Extrinsic membrane proteins are entirely outside of the membrane, but are bound to it by weak molecular attractions (ionic, hydrogen, and/or Van der Waals bonds). Intrinsic membrane proteins, the class we are mainly interested in, are embedded in the membrane. Many of them extend from one side of the membrane to the other and are referred to as transmembrane proteins. Cells are constantly pumping ions in and out through their plasma membranes. In fact, more than half the energy that are bodies consume is used by cells to drive the protein pumps in the brain that do nothing else but transport ions across plasma membranes of nerve cells. How can ions be transported across membranes that are effectively impermeable to them?

Cells contain proteins that are embedded in the lipid bilayer of their plasma membranes and extend from one side of the membrane through to the other. Such transmembrane proteins can function to effect ion transport in several ways. But how can they cope with the energetically highly unfavorable situation in which an ion must pass through the hydrophobic inner layers of the plasma membrane?

2. Domains:

If we examine the detailed structures of many transmembrane proteins, we see that they often have three different domains, two hydrophilic and one hydrophobic. A hydrophilic domain (consisting of hydrophilic amino acids) at the N-terminus is poking out in the extracellular medium, a hydrophobic domain in the middle of the amino acid chain, often only 20-30 amino acids long, is threaded through the plasma membrane, and a hydrophilic domain at the C-terminus protrudes into the cytoplasm. The transmembrane domain, because it is made of amino acids having hydrophobic side chains, exists comfortably in the hydrophobic inner layers of the plasma membrane. Because these transmembrane domains anchor many proteins in the lipid bilayer,these proteins are not free-floating and cannot be isolated and purified biochemically without first dissolving away the lipid bilayer with detergents. (Indeed, much of the washing we do in our lives is necessitated by the need to solubilize proteins that are embedded in lipid membranes using detergents!)

3. Glycoproteins:

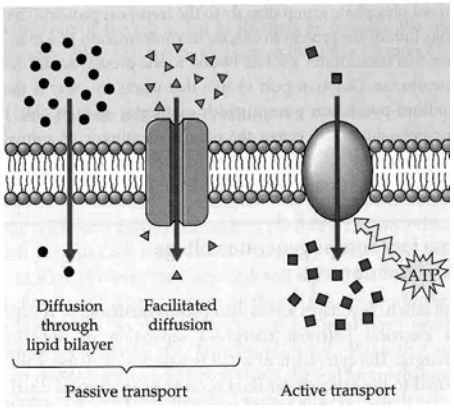
For reasons that are not well understood, many transmembrane proteins are glycoproteins in the sense that sugar side chains are covalently attached to their hydrophilic domains that protrude into the extracellular membrane. A typical mammalian cell may have several hundred distinct types of glycoprotein studding its plasma membrane. Each of these glycoproteins will have its extracellular domain glycosylated with a complex branching bush of sugar residues covalently linked to the asparagine side chains. Some glycoproteins may have 2 or 3 asparagine-linked sugar side chains, others may have dozens.

4. Multi-membrane-spanning proteins:

An elaboration of this scheme depicting membrane proteins having single transmembrane domains involves certain membrane proteins that have multiple transmembrane domains. As one scans along the amino acid sequence of these proteins, it becomes apparent that hydrophilic domains (i.e. having hydrophilic amino acids) alternate with hydrophobic domains. The protein chain as a whole when embedded in the plasma membrane actually weaves back and forth between opposite sides of the plasma membrane. Some think such proteins have the configuration of snakes and hence term them serpentine membrane proteins. A commonly used type of structure seen in many hundreds of serpentine transmembrane proteins involves 7 hydrophobic domains inserted into the plasma membrane separated by hydrophilic regions that are looped out alternatively into either the cytoplasm or the extracellular space.

A3. MEMBRANE TRANSPORT MECHANISM & ION PUMPS

All cells acquire the molecules and ions they need from their surrounding extracellular fluid (ECF). There is an unceasing traffic of molecules and ions in and out of the cell through its plasma membrane. Examples: glucose Na^+, Ca^{2+}. In eukaryotic cells, there is also transport in and out of membrane-bounded intracellular compartments such as the nucleus, endoplasmic reticulum and mitochondria. Examples: proteins, mRNA, Ca^{2+}, ATP



If uncharged solutes are small enough and lipid liking they can move down their concentration gradients directly across the lipid bilayer itself by simple diffusion. Examples of such solutes are ethanol, carbon dioxide, and oxygen.

The major difficulty which the membranes pose for some molecule is that the Lipid bilayers are not permeable to: ions such as

- K^+, Na^+, Ca^{2+} (called cations because when subjected to an electric field they migrate toward the cathode [the negatively-charged electrode])
- Cl^-, HCO_3^- (called anions because they migrate toward the anode [the positively-charged electrode])
- small **hydrophilic** molecules like glucose and **macromolecules** like proteins and RNA

But we all know that that cell is intelligent enough to solve the above challenge posed by the membrane. This is how it solves the problem

Passive and active mechanisms move molecules across membrane.

- Passive transport moves molecules across membrane without expenditure of energy by cell; includes diffusion and facilitated transport.
- Active transport uses energy (ATP) to move molecules across a plasma membrane; include active transport, exocytosis, endocytosis, and pinocytosis.

I. Passive Transport

Cell membranes help organisms maintain homeostasis by controlling what substances may enter or leave the cells. Some substances such as water, oxygen, and carbon dioxide, can cross the cell membrane without any input of energy by the cell. The movement of such substances across the membrane is know as passive transport. The cell membrane is said to be selectively permeable.

The difference in the concentration of molecules across a space is called a concentration gradient. If these molecules diffusing across the membrane from an area of high concentration to an area of low concentration were water molecules the process would be called **osmosis**. Water also moves from a low solute concentration to a high solute concentration.

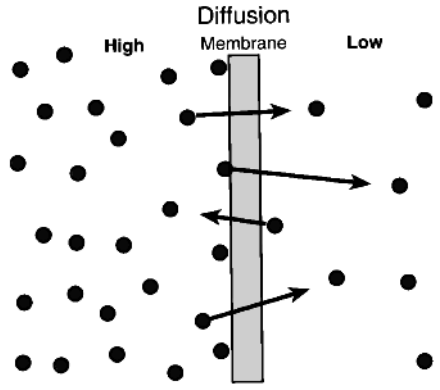


FIGURE The diffusion of gases and lipid-soluble molecules through the lipid bilayer.

The Two types of Passive Transport are

- Transport by simple diffusion and Osmosis
- Facilitated diffusion: carrier proteins and ion channels

1. Diffusion and Osmosis

1. In diffusion, molecules move from higher to lower concentration (i.e., down their concentration gradient).
   - A solution contains a solute, usually a solid, and a solvent, usually a liquid.
   - In the case of a dye diffusing in water, dye is a solute and water is the solvent.
2. Membrane chemical and physical properties allow only a few types of molecules to across by diffusion.
   - Lipid-soluble molecules (e.g., alcohols) diffuse; lipids are membrane's main structural components.
   - Gases readily diffuse through lipid bilayer. Movement of oxygen from air sacs (alveoli) to blood in lung capillaries depends on concentration of oxygen in alveoli.
3. **Osmosis** is the diffusion of water across a differentially permeable membrane from its higher concentration to lower concentration. Osmotic pressure is hydrostatic pressure, on side of membrane with higher solute concentration, produced by water diffusing to that side of membrane
4. Tonicity is strength of a solution in relationship to osmosis; determines movement of water into or out of cells.
   - Isotonic is where the relative solute concentration of two solutions are equal.
   - Hypotonic is where a relative solute concentration of one solution is less than another solution.
   - Hypertonic is where relative solute concentration of one solution is greater than another solution.

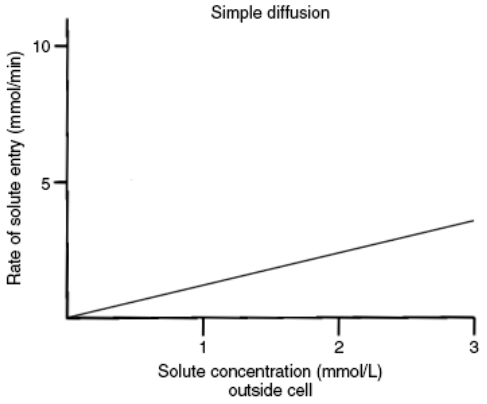


FIGURE **A graph of solute transport across a plasma membrane by simple diffusion.** The rate of solute entry increases linearly with extracellular concentration of the solute. Assuming no change in intracellular concentration, increasing the extracellular concentration increases the gradient that drives solute entry.

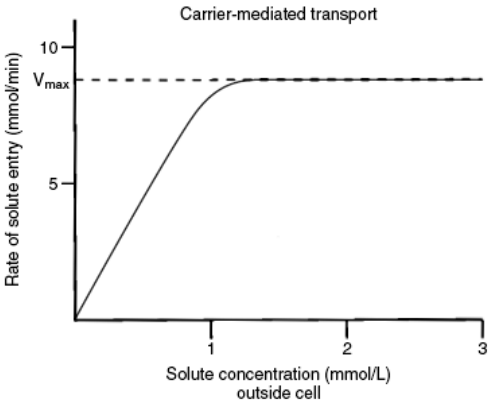


FIGURE **A graph of solute transport across a plasma membrane by carrier-mediated transport.** The rate of transport is much faster than that of simple diffusion and increases linearly as the extracellular solute concentration increases. The increase in transport is limited, however, by the availability of carriers. Once all are occupied by solute, further increases in extracellular concentration have no effect on the rate of transport. A maximum rate of transport (V_{max}) is achieved that cannot be exceeded.

2. Facilitated Diffusion of Ions

Facilitated diffusion of ions takes place through proteins, or assemblies of proteins, embedded in the plasma membrane. These transmembrane proteins form a water-filled channel through which the ion can pass down its concentration gradient

The transmembrane channels that permit facilitated diffusion can be opened or closed. They are said to be "gated". Some types of gated ion channels:

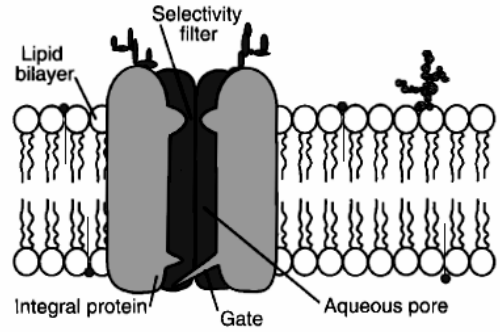


FIGURE **An ion channel.** Ion channels are formed between the polypeptide subunits of integral proteins that span the plasma membrane, providing an aqueous pore through which ions can cross the membrane. Different types of gating mechanisms are used to open and close channels. Ion channels are often selective for a specific ion.

2.1 Ligand-gated Ion Channels: Many ion channels open or close in response to binding a small signaling molecule or "ligand". Some ion channels are gated by extracellular ligands; some by intracellular ligands. In both cases, the ligand is not the substance that is transported when the channel opens.

External ligands: External ligands bind to a site on the extracellular side of the channel.

Examples:

- **Acetylcholine** (ACh). The binding of the **neurotransmitter** acetylcholine at certain **synapses** opens channels that admit Na^+ and initiate a **nerve impulse** or **muscle contraction**.
- **Gamma amino butyric acid (GABA)**: Binding of GABA at certain synapses - designated GABAA - in the central nervous system admits Cl^- ions into the cell and inhibits the creation of a nerve impulse.

Internal ligands: Internal ligands bind to a site on the channel protein exposed to the **cytosol**.

Examples:

- "Second messengers", like **cyclic AMP** (cAMP) and cyclic GMP (cGMP), regulate channels involved in the initiation of impulses in neurons responding to **odors** and **light** respectively.
- ATP is needed to open the channel that allows chloride (Cl^-) and bicarbonate (HCO_3^-) ions out of the cell. This channel is defective in patients with **cystic fibrosis**. Although the energy liberated by the hydrolysis of ATP is needed to open the channel, this is not an example of active transport; the ions diffuse through the open channel following their concentration gradient.

2.2 Mechanically-gated ion Channels

Examples:

- Sound waves bending the cilia-like projections on the hair cells of the inner ear open up ion channels leading to the creation of nerve impulses that the brain interprets as sound.
- Mechanical deformation of the cells of stretch receptors opens ion channels leading to the creation of nerve impulses.

2.3 Voltage-gated ion Channels

In so-called **"excitable" cells** like **neurons** and **muscle cells**, some channels open or close in response to changes in the charge (measured in volts) across the plasma membrane.

Example: As an impulse passes down a neuron, the reduction in the voltage opens sodium channels in the adjacent portion of the membrane. This allows the influx of Na+ into the neuron and thus the continuation of the nerve impulse.

3. Facilitated Diffusion of Molecules

Some small, **hydrophilic** organic molecules, like sugars, can pass through cell membranes by facilitated diffusion. Once again, the process requires transmembrane proteins. In some cases, these - like ion channels -form water-filled pores that enable the molecule to pass in (or out) of the membrane following its concentration gradient.

Examples:

- Maltoporin. This homotrimer in the outer membrane of **E. coli** forms pores that allow the **disaccharide** maltose and a few related molecules to diffuse into the cell.
- The plasma membrane of human red blood cells contain transmembrane proteins that permit the diffusion of glucose from the blood into the cell.
- Glucose transporters- 5 different GLUT proteins and 2 types that cotransport Na and glucose (these are used for secondary active transport)
- Water channels- 8 different types of aquaporins transport water across membrane.

Note that in all cases of facilitated diffusion through channels, the channels are selective; that is, the structure of the protein admits only certain types of molecules through. Whether all cases of facilitated diffusion of small molecules use channels is yet to be proven. Perhaps some molecules are passed through the membrane by a conformational change in the shape of the transmembrane protein when it binds the molecule to be transported.

II. Active Transport: Active transport is the pumping of molecules or ions through a membrane against their concentration gradient. It requires:

- A transmembrane protein (usually a complex of them) called a transporter and energy.
- The source of this energy is **ATP**. The energy of ATP may be used directly or indirectly.

1. Direct Active Transport. Some transporters bind ATP directly and use the energy of its hydrolysis to drive active transport.
2. Indirect Active Transport. Other transporters use the energy already stored in the gradient of a directly-pumped **ion**. Direct active transport of the ion establishes a concentration gradient. When this is relieved by facilitated diffusion, the energy released can be harnessed to the pumping of some other ion or molecule.

Direct Active Transport

1. The Na+/K+ ATPase: The **cytosol** of animal cells contains a concentration of potassium ions (K+) as much as 20 times higher than that in the extracellular fluid. Conversely, the extracellular fluid contains a concentration of sodium ions (Na+) as much as 10 times greater than that within the cell. These concentration gradients are established by the active transport of both ions. And, in fact, the same transporter, called the Na^+/K^+ ATPase, does both jobs. It uses the energy from the hydrolysis of ATP to actively transport 3 Na+ ions out of the cell for each 2 K+ ions pumped into the cell.

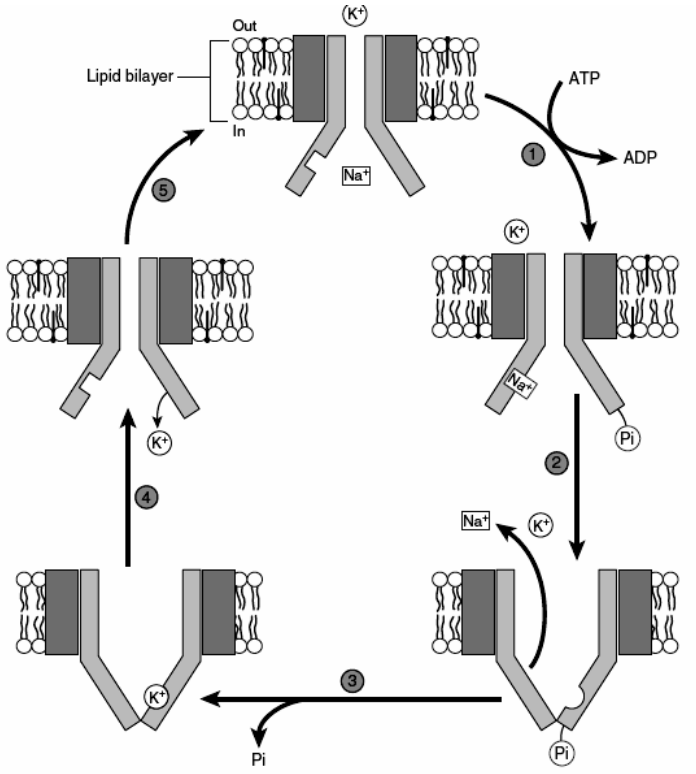


FIGURE **The possible sequence of events during one cycle of the sodium-potassium pump.** The functional form may be a tetramer of two large catalytic subunits and two smaller subunits of unknown function. Binding of intracellular Na^+ and phosphorylation by ATP inside the cell may induce a conformational change that transfers Na^+ to the outside of the cell (steps 1 and 2). Subsequent binding of extracellular K^+ and dephosphorylation return the protein to its original form and transfer K^+ into the cell (steps 3, 4, and 5). There are thought to be three Na^+ binding sites and two K^+ binding sites. During one cycle, three Na^+ are exchanged for two K^+, and one ATP molecule is hydrolyzed.

This accomplishes several vital functions:

- It helps establish a net charge across the plasma membrane with the interior of the cell being negatively charged with respect to the exterior. This **resting potential** prepares nerve and muscle cells for the propagation of action potentials leading to nerve impulses and muscle contraction.
- The accumulation of sodium ions outside of the cell draws water out of the cell and thus enables it to maintain **osmotic balance** (otherwise it would swell and burst from the inward diffusion of water).
- The gradient of sodium ions is harnessed to provide the energy to run several types of **indirect pumps**.

The crucial roles of the Na+/K+ ATPase are reflected in the fact that almost one-third of all the energy generated by the mitochondria in animal cells is used just to run this pump.

2. The H^+/K^+ ATPase: The **parietal cells** of your stomach use this pump to secrete gastric juice. These cells transport protons (H^+) from a concentration of about 4×10^{-8} M within the cell to a concentration of about 0.15 M in the gastric juice (giving it a pH close to 1). Small wonder that parietal cells are stuffed with mitochondria and uses huge amounts of energy as they carry out this three-million fold concentration of protons.

3. The Ca^{2+} ATPase of skeletal muscle : In resting skeletal muscle, there is a much higher concentration of calcium ions (Ca^{2+}) in the sarcoplasmic reticulum than in the cytosol. Activation of the muscle fiber allows some of this Ca^{2+} to pass by facilitated diffusion into the cytosol where it triggers contraction. After contraction, this Ca^{2+} is pumped back into the sarcoplasmic reticulum. This is done by a Ca^{2+} ATPase that uses the energy from each molecule of ATP to pump 2 Ca^{2+} (exchanging them for 2 or 3 protons, H^+).

Pumps **1 - 3** are designated **P-type ion transporters** because they use the same basic mechanism: a conformational change in the proteins as they are reversibly phosphorylated by ATP. And all three pumps can be made to run backward. That is, if the pumped ions are allowed to diffuse back through the membrane complex, ATP can be synthesized from ADP and inorganic phosphate.

4. ABC Transporters

ABC ("**A**TP-**B**inding **C**assette") transporters are **transmembrane proteins** that

- Expose a ligand-binding domain at one surface and a
- ATP-binding domain at the other surface.

The ligand-binding domain is usually restricted to a single type of molecule.The **ATP** bound to its domain provides the energy to pump the ligand across the membrane.The **human genome** contains 48 genes for ABC transporters. **Some examples:**

• **CFTR** - the cystic fibrosis transmembrane conductance regulator
• **TAP**, the transporter associated with antigen processing.
• The transporter that liver cells use to pump the **salts of bile acids** out into the bile.
• ABC transporters that pump chemotherapeutic drugs out of cancer cells thus reducing their effectiveness.

ABC transporters must have evolved early in the history of life. The ATP-binding domains in **archaea**, **eubacteria**, and eukaryotes all share a **homologous** structure, the ATP-binding "cassette".

2. Secondary (Indirect) Active Transport (Co-transport by Symport and Antiport Systems): Indirect active transport also termed as co-transport uses the downhill flow of an ion to pump some other molecule or ion against its gradient. The driving ion is usually sodium (Na^+) with its gradient established by the Na^+/K^+ ATPase.

Besides ATP-powered pumps, cells have a second, discrete class of proteins that import or export ions and small molecules, such as glucose and amino acids, against a concentration gradient. These proteins use the energy stored in the electrochemical gradient of Na+ or H+ ions to power the uphill movement of another substance, a process called **cotransport**.

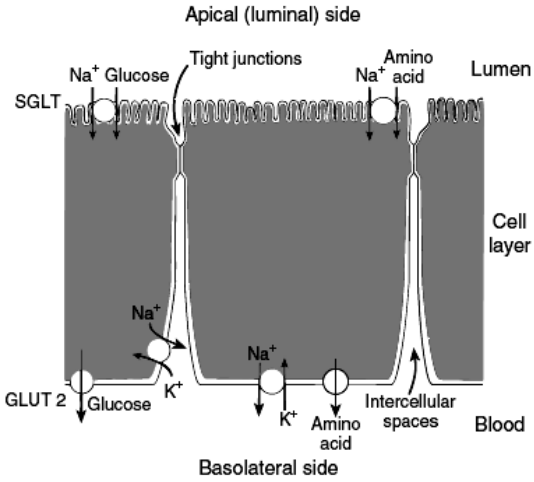


FIGURE **The localization of transport systems to different regions of the plasma membrane in epithelial cells of the small intestine.** A polarized cell is produced, in which entry and exit of solutes, such as glucose, amino acids, and Na^+, occur at opposite sides of the cell. Active entry of glucose and amino acids is restricted to the apical membrane and exit requires equilibrating carriers located only in the basolateral membrane. For example, glucose enters on SGLT and exits on GLUT 2. Na^+ that enters via the apical symporters is pumped out by the Na^+/K^+-ATPase on the basolateral membrane. The result is a net movement of solutes from the luminal side of the cell to the basolateral side, ensuring efficient absorption of glucose, amino acids, and Na^+ from the intestinal lumen.

1. Symport: In this type of indirect active transport, the driving ion (Na+) and the pumped molecule pass through the membrane pump in the **same** direction.

Example 1: Na^+-Linked Symporters Import Amino Acids and Glucose into Many Animal Cells. Many cells, such as those lining the small intestine and the kidney tubules, need to concentrate glucose against a very large concentration gradient. Such cells utilize a *two Na+/one-glucose* **symporter**; a protein that couples transmembrane movement of one glucose molecule to the transport of two Na+ ions. The Na+/glucose transporter is used to actively transport glucose out of the intestine and also out of the **kidney tubules** and back into the blood.

Example 2:The Na+/iodide transporter. This symporter pumps iodide ions into the cells of the **thyroid gland** (for the manufacture of thyroxine) and also into the cells of the mammary gland (to supply the baby's need for iodide).

2. Antiport: Cell uses movement of an ion across a membrane and down its concentration gradient to power the transport of a second substance "uphill" against its gradient. In this process, the two substances move across the membrane in opposite directions.

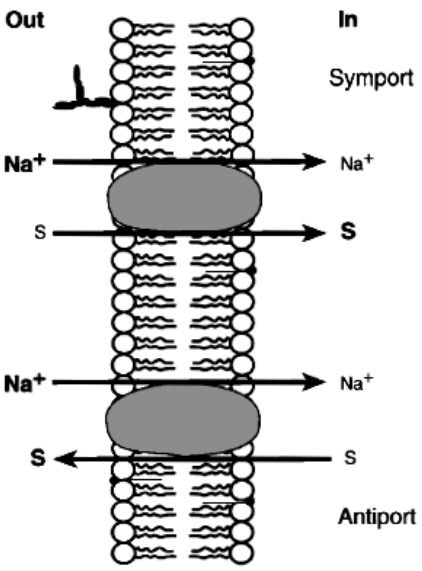


FIGURE **Secondary active transport systems.** In a symport system (top), the transported solute (S) is moved in the same direction as the Na^+ ion. In an antiport system (bottom), the solute is moved in the opposite direction to Na^+. Large and small type indicate high and low concentrations, respectively, of Na^+ ions and solute.

Example 1: Na^+-Linked Antiporter Exports Ca^{2+} from Cardiac Muscle Cells

- It is one of the best cation antiporter system. The plasma membrane of most cells contains one or more types of antiporters, which couple movement of a co transported ion (often Na+) down its electrochemical gradient to movement of a different molecule in the opposite direction against a concentration gradient .In cardiac muscle cells, for example, a *Na+/Ca*2+ *antiporter,* rather than a plasma membrane Ca2+ ATPase, plays the principal role in maintaining a low concentration of Ca2+ in the cytosol.
- It will be interesting to see how the following two drugs work. The drugs ***ouabain*** **and** ***digoxin*** increase the force of heart muscle contraction and are widely used in the treatment of congestive heart failure. The primary effect of these drugs is to inhibit the Na+/K+ ATPase, thereby raising the intracellular Na+ concentration (and lowering intracellular K+). Because the Na+/Ca2+ antiporter functions less efficiently with a lower Na+ concentration gradient, fewer Ca2+ ions are exported and the intracellular Ca2+ concentration increases. This increase causes the muscle to contract more strongly

Example 2: AE1 Protein, a Cl^- /HCO_3^- Antiporter, Is Crucial to CO2 Transport by Erythrocytes

- In addition to cation antiporters, which transport only positive ions, many cells also contain *anion transporters*, which transport only negative ions. An important example is *AE1 protein*, the predominant integral protein of the mammalian erythrocyte.
- This anion antiporter catalyzes the one-for-one exchange of Cl and HCO3 across the plasma membrane. Since one singly charged negative ion is exchanged for another, there is no net movement of electric charge and the reaction is not affected by the membrane potential. Thus, the direction of the reaction is dependent only on the concentration gradients of the transported ions.
- Transmembrane anion exchange is essential to an important function of the erythrocyte the transport of waste carbon dioxide (CO2), which is generated in peripheral tissues, to the lungs for excretion by respiratory exhalation Waste CO2 released from cells into the capillary blood diffuses across the erythrocyte membrane. In its gaseous form, CO_2 dissolves poorly in aqueous solutions, such as the cytosol or blood plasma, but the potent enzyme *carbonic anhydrase* inside the erythrocyte converts CO_2 to the water-soluble bicarbonate (HCO_3) anion. The release of oxygen from hemoglobin int`o the peripheral capillaries induces a conformational change in the globin polypeptide that enables a histidine side chain to bind the proton produced by the carbonic anhydrase reaction. Meanwhile, the HCO_3 formed by carbonic anhydrase is transported out of the erythrocyte in exchange for an entering Cl via AE1 protein
- The overall direction of this anion-exchange process is reversed in the lungs. CO_2 diffuses out of the erythrocyte and is eventually expelled in breathing. The lowered concentration of CO_2 within the cytosol drives the carbonic anhydrase reaction, as written above, from right to left: HCO_3 reacts to yield CO_2 and OH. At the same time, oxygen binding to hemoglobin causes a proton to be released from hemoglobin; the proton combines with the OH to form H_2O. The lowered intracellular HCO_3 concentration causes HCO_3 to enter the erythrocyte in exchange for Cl

Example 3: Several Cotransporters Regulate Cytosolic pH

- The anaerobic metabolism of glucose yields lactic acid, and aerobic metabolism yields CO_2, which is hydrated by carbonic anhydrase to carbonic acid (H_2CO_3). These weak acids dissociate, yielding H+ ions (protons); if these protons were not exported from cells, the cytosolic pH would drop precipitously, endangering cellular functioning. Two types of cotransport proteins are employed to remove some of the "excess" protons generated during metabolism of animal cells.
- One is a *Na^+HCO_3 /Cl cotransporter,* which imports one Na^+ ion down its concentration gradient, together with one HCO_3 , in exchange for export of one Cl ion against its concentration gradient. The imported HCO_3 ions combine with protons generated by metabolism to produce CO_2, which diffuses out of the cell. Thus the overall action of this transporter raises the cytosolic pH (reduces the H+ concentration). Also important in removing excess protons is a *Na^+/H^+ antiporter,* which couples entry of one Na+ ion into the cell down its concentration gradient to export one H^+ ion.
- The plasma membranes of most animal cells also contain a *Na+-independent Cl /HCO_3 antiporter* similar to the erythrocyte AE1 protein discussed previously. This anion-exchange protein functions to lower the cytosolic pH, in effect removing "excess" OH ions. Recall that a HCO_3 ion can be viewed as a complex of OH and CO_2, so export of HCO_3 lowers the cytosolic pH. Exchange of cytosolic HCO_3 for extracellular Cl is powered by the import of Cl down its concentration gradient (Cl out > Cl in).
- The activity of all three of these antiport proteins depends on pH, providing cells with a fine-tuned mechanism for controlling the cytosolic pH The protonexporting transporters, which are activated when the pH of the cytosol falls, act to raise the cytosolic pH. Similarly, a rise in pH above 7

stimulates the Cl / HCO_3 antiporter, leading to a more rapid export of HCO_3 and decrease in the cytosolic pH. In this manner the cytosolic pH of growing cells is maintained very close to pH 7.4.

- Small changes in the cytosolic pH may have profound effects on the overall cellular metabolic rate. For instance, primary fibroblast cells grown to maximal density (confluence) in tissue culture generally become quiescent: DNA synthesis stops; the rates of RNA synthesis, glucose catabolism, and protein synthesis are reduced; and the cytosolic pH drops from the characteristic 7.4 of growing cells to pH 7.2. Treatment of quiescent cells with a mixture of serum growth factors restimulates cell growth and DNA synthesis. An early effect of these growth factors is a marked increase in the cytosolic pH to 7.4; this dramatic change is caused in part by stimulation of the Na^+/H^+ antiport, which expels protons into the medium. The rise in cytosolic pH is believed to help activate certain metabolic pathways required for cell growth and division.

Example 4: Numerous Transport Proteins Enable Plant Vacuoles to Accumulate Metabolites and Ions

- The lumen of plant vacuoles is much more acidic (pH 3 to 6) than is the cytosol (pH 7.5). As noted earlier, the vacuolar membrane contains a V-class ATP-powered pump and a unique PPi-powered pump, both of which function to pump H+ ions into the vacuolar lumen against a concentration gradient. The vacuolar membrane also contains Cl and NO3 channels that transport these anions from the cytosol into the vacuole. Entry of these anions against their concentration gradients is driven by the inside-positive potential generated by the H+ pumps. Operation of both types of proton pumps in conjunction with these anion channels produces an inside-positive electric potential of about 20 mV across the vacuolar membrane and also a substantial pH gradient.
- The proton gradient and electric potential across the plant vacuole membrane are used in much the same way as the Na+ gradient and electric potential across the animal-cell plasma membrane: to power the selective uptake or extrusion of ions and small molecules. In the leaf, for example, excess sucrose generated during photosynthesis in the day is stored in the vacuole; during the night the stored sucrose moves into the cytoplasm and is metabolized to CO2 and H2O with concomitant generation of ATP from ADP and Pi.
- A *protonsucrose antiporter* in the vacuolar membrane operates to accumulate sucrose in plant vacuoles. The inward movement of sucrose is powered by the outward movement of H+, which is favored by its concentration gradient (lumen > cytosol) and by the outward-negative potential across the vacuolar membrane Uptake of Ca2+ and Na+ into the vacuole from the cytosol against their concentration gradients is similarly mediated by proton antiporters.

The **Na^+/K^+ ATPase** is also an antiport pump using the energy of ATP to pump Na+ out of the cell; K+ in.An antiport pump in the vacuole of some plants harnesses the outward facilitated diffusion of protons (themselves pumped into the vacuole by a **H^+ ATPase**) to the active inward transport of sodium ions. This sodium/proton Antiport pump enables the plant to sequester sodium ions in its vacuole.

III. Movement of Large Molecules: Large molecules (proteins, nucleic acids, polypeptides larger than a few amino acids, polysaccharides larger than a few sugars) are not carried by transport proteins. There are mechanisms for moving larger molecules, but they don't enter into cytoplasm.

1. Endocytosis, is a general term for the process whereby very large particles of material are wrapped with plasma membrane and moved into the cell in the form of vesicles or vacuoles. None of the trapped material actually moves *through* the membrane, but remains on the other side of the original membrane, even while the vacuole is inside the cell. Thus it is defined as a process where a cell takes in macromolecules by forming vesicles derived from the plasma membrane. Vesicle forms from a localized region of plasma membrane that sinks inward; pinches off into the cytoplasm. It is used by cells to incorporate extracellular substances.

There are three types of endocytosis:

a. Pinocytosis ("cell drinking"): It is almost the same process as phagocytosis, except it involves liquids instead of solids. During exocytosis a vacuole containing material to be excreated from the cell moves to the plasma membrane and fuses with it.

- The vacuole membrane becomes part of the plasma membrane and the contents are released to the outside. Cells use this method to eliminate the wastes left after digestion and metabolism and also to release a whole variety of materials that have been synthesized inside the cell but which are needed outside the cell. Release of hormones and digestive enzymes, found in multicellular animals, are two examples of this process.

b. Phagocytosis("cell eating"): Particles are engulfed by phagocytosis a process that begins when solids make contact with the outer cell surface, triggering the movement of the membrane.

- The desired particles are then enclosed within a small piece of the plasma membrane which forms a sac called a vacuole (or vesicle), with the food particle inside it. This vacuole is then moved to the interior of the cell. Strictly speaking, the food particles are not yet part of the cell as it is still surrounded by membrane.
- Before food can be used, it must be broken down to smaller pieces and those pieces moved into the cytoplasm. Digestion occurs when the food vacuole is fused with a second vacuole, called a lysosome, that contains powerful digestive enzymes. Food is degraded, its nutrients are absorbed by the cell and its waste products are left in the digestive vacuole, which may then leave the cell by exocytosis.
- Phagocytosis occurs in the scavenging white blood cells of our body. They prowl around looking for invading bacteria and viruses which they engulf and destroy them

c. Receptor-Mediated Endocytosis: Endocytosis when coated pits form vesicles when specific ligands bind to receptors on the cell's surface. It is More discriminating process than pinocytosis.

- Enables cells to acquire bulk quantities of specific substances, even if they are in low concentration in extracellular fluid (e.g. cholesterol).
- Membrane-embedded proteins with specific receptor sites exposed to the cell's exterior, cluster in regions called coated pits.
- A layer of clathrin, a fibrous protein, lines and reinforces the coated pit on the cytoplasmic side.
- A molecule that binds to a specific receptor site of another molecule is call a ligand.

- Eg. Iron is carried through blood tightly bound to transferrin protein carrier. To get iron into cells, cell membrane contains special receptor proteins that bind transferrin, move towards special regions of membrane under which lie **clathrin** proteins. Endocytosis occurs inside clathrin "cage", moves inside cell. Cage eventually recycles back to cell surface, returning transferrin proteins to cell exterior. However, iron is released inside cell, exits from vesicles, becomes bound to ferritin.

2. Exocytosis is the reverse of endocytosis. Quantities of material are expelled from the cell without ever passing through the membrane as individual molecules.

- By using the processes of endocytosis and exocytosis, some specialized types of cells move large amounts of bulk material into and out of them
- It is also an active transport process.
- During exocytosis a vesicle moves to the cell membrane, fuses with it, and then releases it contents to the outside of the cell.

- Vesicle usually budded from the ER or Golgi and migrates to plasma membrane. It is mainly used by secretory cells to export products (e.g. insulin in pancreas, or neurotransmitter from neuron.

A4. Mechanism of sorting and regulation of intracellular transport

The protein biosynthetic pathways in cells can be considered to be **one large sorting system.** Many proteins carry **signals** (usually but not always specific sequences of amino acids) that direct them to their destination, thus ensuring that they will end up in the appropriate membrane or cell compartment; these signals are a fundamental component of the sorting system. Usually the signal sequences are recognized and interact with complementary areas of proteins that serve as receptors for the proteins that contain them.

A major sorting decision is made early in protein biosynthesis, when specific proteins are synthesized either on free or on membrane-bound polyribosomes. This results in two sorting branches called the **cytosolic branch** and the **rough endoplasmic reticulum (RER) branch** (Figure 1). This sorting occurs because proteins synthesized on membrane-bound polyribosomes contain a **signal peptide** that mediates their attachment to the membrane of the ER. Further details on the signal peptide are given below. Proteins synthesized on **free polyribosomes** lack this particular signal peptide and are delivered into the cytosol. There they are directed to mitochondria, nuclei, and peroxisomes by specific signals—or remain in the cytosol if they lack a signal. Any protein that contains a targeting sequence that is subsequently removed is designated as a **preprotein.**

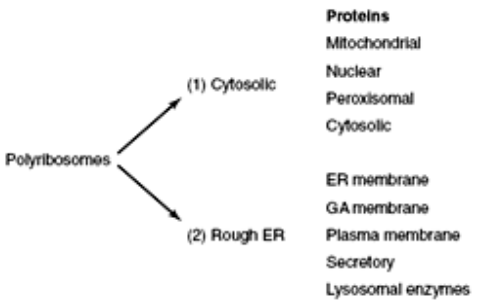


Figure 1. Diagrammatic representation of the two branches of protein sorting occurring by synthesis on (1) cytosolic and (2) membrane-bound polyribosomes. The mitochondrial proteins listed are encoded by nuclear genes. Some of the signals used in further sorting of these proteins are listed in Table 4.

In some cases a second peptide is also removed, and in that event the original protein is known as a **preproprotein** (eg, preproalbumin). Proteins synthesized and sorted in the **rough ER branch** (Figure 2) include many destined for various membranes (eg, of the ER, Golgi apparatus, lysosomes, and plasma membrane) and for secretion. Lysosomal enzymes are also included. Thus, such proteins may reside in the membranes or lumens of the ER or follow the major transport route of intracellular proteins to the Golgi apparatus. Further signal-mediated sorting of certain proteins occurs in the Golgi apparatus, resulting in delivery to lysosomes, membranes of the Golgi apparatus, and other sites. Proteins destined for the plasma membrane or for secretion pass through the Golgi apparatus but generally are not thought to carry specific sorting signals; they are believed to reach their destinations by default.

The entire pathway of ER → Golgi apparatus → plasma membrane is often called the **secretory** or **exocytotic pathway.** Events along this route will be given special attention. Most of the proteins reaching the Golgi apparatus or the plasma membrane are carried in **transport vesicles;** a brief description of the formation of these important particles will be given subsequently.

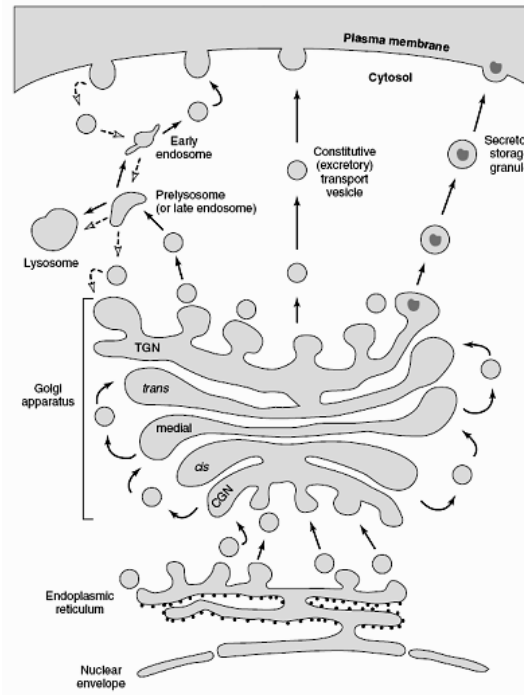


Figure 2. Diagrammatic representation of the rough endoplasmic reticu lum branch of protein sorting.

Other proteins destined for secretion are carried in **secretory vesicles** (Figure 2). These are prominent in the pancreas and certain other glands. Their mobilization and

discharge are regulated and often referred to as **"regulated secretion,"** whereas the secretory pathway involving transport vesicles is called **"constitutive."**

Experimental approaches that have afforded major insights to the processes described in this chapter include (1) use of yeast mutants; (2) application of recombinant DNA techniques (eg, mutating or eliminating particular sequences in proteins, or fusing new sequences onto them; and (3) development of in vitro systems (eg, to study translocation in the ER and mechanisms of vesicle formation). The sorting of proteins belonging to the cytosolic branch referred to above is described next, starting with
mitochondrial proteins.

THE MITOCHONDRION BOTH IMPORTS & SYNTHESIZES PROTEINS

Mitochondria contain many proteins. Thirteen proteins (mostly membrane components of the electron transport chain) are encoded by the **mitochondrial genome** and synthesized in that organelle using its own protein-synthesizing system. However, the majority (at least several hundred) are encoded by **nuclear genes,** are synthesized outside the mitochondria on cytosolic polyribosomes, and must be imported. **Yeast cells** have proved to be a particularly useful system for analyzing the mechanisms of import of mitochondrial proteins, partly because it has proved possible to generate a variety of mutants that have illuminated the fundamental processes involved. Most progress has been made in the study of proteins present in the mitochondrial matrix, such as the F1 ATPase subunits. Only the pathway of import of matrix proteins will be discussed in any detail here.

Matrix proteins must pass from cytosolic polyribosomes through the outer and inner mitochondrial membranes to reach their destination. Passage through the two membranes is called **translocation.** They have an amino terminal leader sequence (presequence), about 20–80 amino acids in length, which is not highly conserved but contains many positively charged amino acids (eg, Lys or Arg). The presequence is equivalent to a signal peptide mediating attachment of polyribosomes to membranes of the ER (see below), but in this instance targeting proteins to the matrix; if the leader sequence is cleaved off, potential matrix proteins will not reach their destination.

Translocation is believed to occur **posttranslationally,** after the matrix proteins are released from the cytosolic polyribosomes. Interactions with a number of cytosolic proteins that act as **chaperones** (see below) and as targeting factors occur prior to translocation. Two distinct **translocation complexes** are situated in the outer and inner mitochondrial membranes, referred to (respectively) as TOM (translocase-of-the outer membrane) and TIM (translocase-of-the-inner membrane). Each complex has been analyzed and found to be composed of a number of proteins, some of which act as receptors for the incoming proteins and others as components of the transmembrane pores through which these proteins must pass. Proteins must be in the **unfolded state** to pass through the complexes, and this is made possible by **ATP-dependent binding to several chaperone proteins.** In mitochondria, they are involved in translocation, sorting, folding, assembly, and degradation of imported proteins. A **proton-motive force** across the inner membrane is required for import; it is made up of the **electric potential** across the membrane (inside negative) and the **pH gradient**. The positively charged leader sequence may be helped through the membrane by the negative charge in the matrix. The presequence is split off in the matrix by a **matrix-processing peptidase (MPP).** Contact with **other chaperones** present in the matrix is essential to complete the overall process of import. Interaction with mt-Hsp70 (Hsp = heat shock protein) ensures proper import into the matrix and prevents misfolding or aggregation, while interaction with the mt-Hsp60- Hsp10 system ensures proper folding. The latter proteins resemble the bacterial GroEL chaperonins, a subclass of chaperones that form complex cage-like assemblies made up of heptameric ring structures. The interactions of imported proteins with the above chaperones require **hydrolysis of ATP** to drive them.

The details of how preproteins are translocated have not been fully elucidated. It is possible that the electric potential associated with the inner mitochondrial membrane causes a conformational change in the unfolded preprotein being translocated and that this helps to pull it across. Furthermore, the fact that the matrix is more negative than the intermembrane space may "attract" the positively charged amino terminal of the preprotein to enter the matrix. Close contact between the membrane sites in the outer and inner membranes involved in translocation is necessary.

The above describes the major pathway of proteins destined for the mitochondrial matrix. However, certain proteins insert into the **outer mitochondrial membrane** facilitated by the TOM complex. Others stop in the **intermembrane space,** and some insert into the **inner membrane.** Yet others proceed into the matrix and then return to the inner membrane or intermembrane space. A number of proteins contain two signaling sequences—one to enter the mitochondrial matrix and the other to mediate subsequent relocation (eg, into the inner membrane). Certain mitochondrial proteins do not contain presequences (eg, cytochrome *c,* which locates in the inter membrane space), and others contain **internal presequences.** Overall, proteins employ a variety of mechanisms and routes to attain their final destinations in mitochondria. General features that apply to the import of proteins into organelles, including mitochondria and some of the other organelles to be discussed below, are summarized in Table 1.

Table 1. Some general features of protein import to organelles.

- Import of a protein into an organelle usually occurs in three stages: recognition, translocation, and maturation.
- Targeting sequences on the protein are recognized in the cytoplasm or on the surface of the organelle.
- The protein is unfolded for translocation, a state maintained in the cytoplasm by chaperones.
- Threading of the protein through a membrane requires energy and organellar chaperones on the trans side of the membrane.
- Cycles of binding and release of the protein to the chaperone result in pulling of its polypeptide chain through the membrane.
- Other proteins within the organelle catalyze folding of the protein, often attaching cofactors or oligosaccharides and assembling them into active monomers or oligomers.

IMPORTINS & EXPORTINS ARE INVOLVED IN TRANSPORT OF MACROMOLECULES IN & OUT OF THE NUCLEUS

It has been estimated that more than a million macromolecules per minute are transported between the nucleus and the cytoplasm in an active eukaryotic cell. These macromolecules include histones, ribosomal proteins and ribosomal subunits, transcription factors, and mRNA molecules. The transport is bidirectional and occurs through the nuclear pore complexes (NPCs).

These are complex structures with a mass approximately 30 times that of a ribosome and are composed of about 100 different proteins. The diameter of an NPC is approximately 9 nm but can increase up to approximately 28 nm. Molecules smaller than about 40 kDa can pass through the channel of the NPC by diffusion, but special translocation mechanisms exist for larger molecules. These mechanisms are under intensive investigation, but some important features have already emerged.

Here we shall mainly describe **nuclear import** of certain macromolecules. The general picture that has emerged is that proteins to be imported (cargo molecules) carry a **nuclear localization signal (NLS).** One example of an NLS is the amino acid sequence $(Pro)_2$- $(Lys)_4$-Ala-Lys-Val, which is markedly rich in basic lysine residues. Depending on which NLS it contains, a cargo molecule interacts with one of a family of soluble proteins called **importins,** and the complex **docks** at the NPC. Another family of proteins called **Ran** plays a critical regulatory role in the interaction of the complex with the NPC and in its translocation through the NPC. Ran proteins are small monomeric nuclear GTPases and, like other GTPases, exist in either GTPbound or GDP-bound states. They are themselves regulated by **guanine nucleotide exchange factors** (GEFs; eg, the protein RCC1 in eukaryotes), which are located in the nucleus, and Ran **guanine-activating proteins** (GAPs), which are predominantly cytoplasmic. The GTP-bound state of Ran is favored in the nucleus and the GDP-bound state in the cytoplasm. The conformations and activities of Ran molecules vary depending on whether GTP or GDP is bound to them (the GTP-bound state is active). The **asymmetry** between nucleus and cytoplasm—with respect to which of these two nucleotides is bound to Ran molecules—is thought to be crucial in understanding the roles of Ran in transferring complexes unidirectionally across the NPC. When cargo molecules are released inside the nucleus, the importins recirculate to the cytoplasm to be used again. Figure 3 summarizes some of the principal features in the above process.

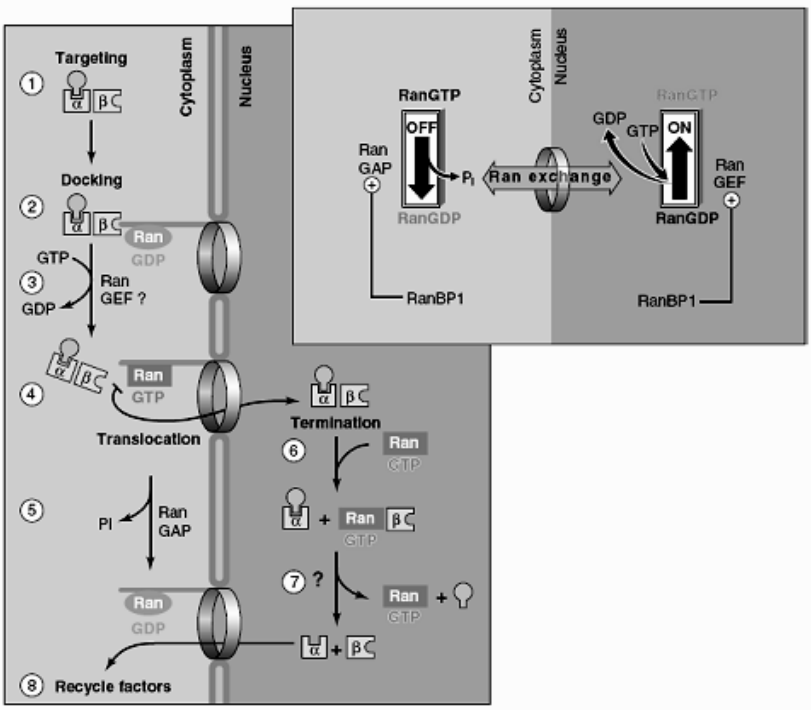


Figure 3. Schematic representation of the proposed role of Ran in the import of cargo carrying an NLS signal. (1) The targeting complex forms when the NLS receptor (α, an importin) binds NLS cargo and the docking factor (β). (2) Docking occurs at filamentous sites that protrude from the NPC. Ran-GDP docks independently. (3) Transfer to the translocation channel is triggered when a RanGEF converts Ran-GDP to Ran-GTP. (4) The NPC catalyzes translocation of the targeting complex. (5) Ran-GTP is recycled to Ran-GDP by docked RanGAP. (6) Ran-GTP disrupts the targeting complex by binding to a site on β that overlaps with a binding site. (7) NLS cargo dissociates from α, and Ran-GTP may dissociate from β. (8) α and β factors are recycled to the cytoplasm. *Inset:* The Ran translocation switch is off in the cytoplasm and on in the nucleus. Ran-GTP promotes NLS- and NES-directed translocation. However, cytoplasmic Ran is enriched in Ran-GDP (OFF) by an active RanGAP, and nuclear pools are enriched in Ran-GTP (ON) by an active GEF. RanBP1 promotes the contrary activities of these two factors. Direct linkage of nuclear and cytoplasmic pools of Ran occurs through the NPC by an unknown shuttling mechanism. P_i, inorganic phosphate; NLS, nuclear localization signal; NPC, nuclear pore complex; GEF, guanine nucleotide exchange factor; GAP, guanine-activating protein; NES, nuclear export signal; BP, binding protein.

Other **small monomeric GTPases** (eg, ARF, Rab, Ras, and Rho) are important in various cellular processes such as vesicle formation and transport (ARF and Rab; see below), certain growth and differentiation processes (Ras), and formation of the actin cytoskeleton. A process involving GTP and GDP is also crucial in the transport of proteins across the membrane of the ER (see below).

Proteins similar to importins, referred to as **exportins,** are involved in export of many macromolecules from the

nucleus. Cargo molecules for export carry **nuclear export signals (NESs).** Ran proteins are involved in this process also, and it is now established that the processes of import and export share a number of common features.

MOST CASES OF ZELLWEGER SYNDROME ARE DUE TO MUTATIONS IN GENES INVOLVED IN THE BIOGENESIS OF PEROXISOMES

The peroxisome is an important organelle involved in aspects of the metabolism of many molecules, including fatty acids and other lipids (eg, plasmalogens, cholesterol, bile acids), purines, amino acids, and hydrogen peroxide. The peroxisome is bounded by a single membrane and contains more than 50 enzymes; catalase and urate oxidase are marker enzymes for this organelle. Its proteins are synthesized on cytosolic polyribosomes and fold prior to import. The pathways of import of a number of its proteins and enzymes have been studied, some being **matrix components** and others **membrane components.**

At least two **peroxisomal-matrix targeting sequences (PTSs)** have been discovered. One, PTS1, is a tripeptide (ie, Ser-Lys-Leu [SKL], but variations of this sequence have been detected) located at the carboxyl terminal of a number of matrix proteins, including catalase. Another, PTS2, consisting of about 26–36 amino acids, has been found in at least four matrix proteins (eg, thiolase) and, unlike PTS1, is cleaved after entry into the matrix. Proteins containing PTS1 sequences form complexes with a soluble receptor protein (PTS1R) and proteins containing PTS2 sequences complex with another, PTS2R. The resulting complexes then interact with a membrane receptor, Pex14p. Proteins involved in further transport of proteins into the matrix are also present. Most peroxisomal membrane proteins have been found to contain neither of the above two targeting sequences, but apparently contain others. The import system can handle **intact oligomers** (eg, tetrameric catalase). Import of matrix proteins requires ATP, whereas import of membrane proteins does not.

Interest in import of proteins into peroxisomes has been stimulated by studies on **Zellweger syndrome.** This condition is apparent at birth and is characterized by profound neurologic impairment, victims often dying within a year. The number of peroxisomes can vary from being almost normal to being virtually absent in some patients. Biochemical findings include an accumulation of very long chain fatty acids, abnormalities of the synthesis of bile acids, and a marked reduction of plasmalogens. The condition is believed to be due to mutations in genes encoding certain proteins—so called **peroxins**—involved in various steps of **peroxisome biogenesis** (such as the import of proteins described above), or in genes encoding certain peroxisomal enzymes themselves. Two closely related conditions are **neonatal adrenoleukodystrophy** and **infantile Refsum disease.** Zellweger syndrome and these two conditions represent a spectrum of overlapping features, with Zellweger syndrome being the most severe (many proteins affected) and infantile Refsum disease the least severe (only one or a few proteins affected).

THE SIGNAL HYPOTHESIS EXPLAINS HOW POLYRIBOSOMES BIND TO THE ENDOPLASMIC RETICULUM

As indicated above, the rough ER branch is the second of the two branches involved in the synthesis and sorting of proteins. In this branch, proteins are synthesized on membrane-bound polyribosomes and translocated into the lumen of the rough ER prior to further sorting (Figure 2).

The **signal hypothesis** was proposed by Blobel and Sabatini partly to explain the distinction between free and membrane-bound polyribosomes. They found that proteins synthesized on membrane-bound polyribosomes contained a peptide extension **(signal peptide)** at their amino terminals which mediated their attachment to the membranes of the ER. As noted above, proteins whose entire synthesis occurs on free polyribosomes lack this signal peptide. An important aspect of the signal hypothesis was that it suggested—as turns out to be the case—that **all ribosomes have the same structure** and that the distinction between membranebound and free ribosomes depends solely on the former's carrying proteins that have signal peptides. Much evidence has confirmed the original hypothesis. Because many membrane proteins are synthesized on membrane- bound polyribosomes, the signal hypothesis plays an important role in concepts of membrane assembly. Some characteristics of signal peptides are summarized in Table 3.

Table 3. Some properties of signal peptides.

- Usually, but not always, located at the amino terminal
- Contain approximately 12–35 amino acids
- Methionine is usually the amino terminal amino acid
- Contain a central cluster of hydrophobic amino acids
- Contain at least one positively charged amino acid near their amino terminal
- Usually cleaved off at the carboxyl terminal end of an Ala residue by signal peptidase

Figure 4 illustrates the principal features in relation to the passage of a secreted protein through the membrane of the ER. It incorporates features from the original signal hypothesis and from subsequent work.

The mRNA for such a protein encodes an amino terminal **signal peptide** (also variously called a leader sequence, a transient insertion signal, a signal sequence, or a presequence). The signal hypothesis proposed that the protein is inserted into the ER membrane at the same time as its mRNA is being translated on polyribosomes, so-called **cotranslational insertion.** As the signal peptide emerges from the large subunit of the ribosome, it is recognized by a **signal recognition particle (SRP)** that blocks further translation after about 70 amino acids have been polymerized (40 buried in the large ribosomal subunit and 30 exposed). The block is referred to as **elongation arrest.** The SRP contains six proteins and has a 7S RNA associated with it that is closely related to the Alu family of highly repeated DNA sequences. The SRP-imposed block is not released until the SRP-signal peptide-polyribosome complex has bound to the so-called **docking protein** (SRP-R, a receptor for the SRP) on the ER membrane; the SRP thus guides the signal peptide to the SRP-R and prevents premature folding and expulsion of the protein being synthesized into the cytosol. The SRP-R is an integral membrane protein composed of α and β subunits. The α subunit binds GDP and the β subunit spans the membrane. When the SRP signal peptide complex interacts with the receptor, the exchange of GDP for GTP is stimulated. This form of the receptor (with GTP bound) has a high affinity for the SRP and thus releases the signal peptide, which binds to the translocation machinery (translocon) also present in the ER membrane. The α subunit then hydrolyzes its bound GTP, restoring GDP and completing a GTP-GDP cycle. The unidirectionality of this cycle helps drive the interaction of the polyribosome and its signal peptide with the ER membrane in the forward direction. The **translocon** consists of a number of membrane proteins that form a protein-conducting channel in the ER membrane through which the newly synthesized protein may pass. The channel appears to be open only when a signal peptide is present, preserving conductance across the ER membrane when it closes. The conductance of the channel has been measured experimentally.

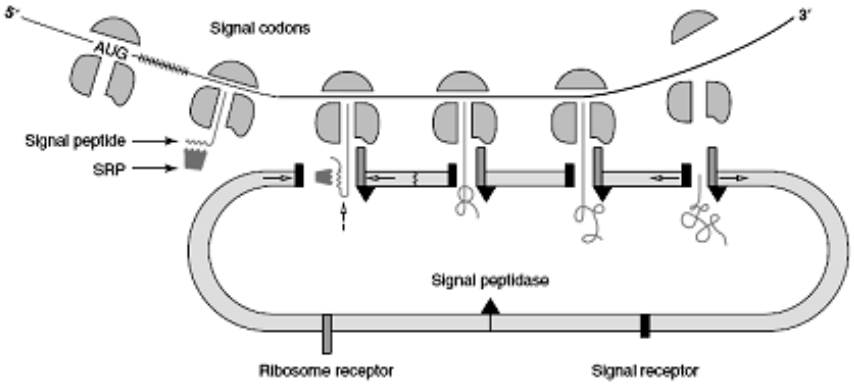


Figure 4. Diagram of the signal hypothesis for the transport of secreted proteins across the ER membrane. The ribosomes synthesizing a protein move along the messenger RNA specifying the amino acid sequence of the protein. (The messenger is represented by the line between 5′ and 3′.) The codon AUG marks the start of the message for the protein; the hatched lines that follow AUG represent the codons for the signal sequence. As the protein grows out from the larger ribosomal subunit, the signal sequence is exposed and bound by the signal recognition particle (SRP). Translation is blocked until the complex binds to the "docking protein," also designated SRP-R (represented by the solid bar) on the ER membrane. There is also a receptor (open bar) for the ribosome itself. The interaction of the ribosome and growing peptide chain with the ER membrane results in the opening of a channel through which the protein is transported to the interior space of the ER. During translocation, the signal sequence of most proteins is removed by an enzyme called the "signal peptidase," located at the luminal surface of the ER membrane. The completed protein is eventually released by the ribosome, which then separates into its two components, the large and small ribosomal subunits. The protein ends up inside the ER. See text for further details.

Specific functions of a number of components of the translocon have been identified or suggested. **TRAM** (translocating chain-associated membrane) protein may bind the signal sequence as it initially interacts with the translocon and the **Sec61p** complex (consisting of three proteins) binds the heavy subunit of the ribosome. The insertion of the signal peptide into the conducting channel, while the other end of the parent protein is still attached to ribosomes, is termed **"cotranslational insertion."** The process of elongation of the remaining portion of the protein probably facilitates passage of the nascent protein across the lipid bilayer as the ribosomes remain attached to the membrane of the ER. Thus, the rough (or ribosome-studded) ER is formed. It is important that the protein be kept in an **unfolded state** prior to entering the conducting channel—otherwise, it may not be able to gain access to the channel.

Ribosomes remain attached to the ER during synthesis of signal peptide-containing proteins but are released and dissociated into their two types of subunits when the process is completed. The signal peptide is hydrolyzed by **signal peptidase,** located on the luminal side of the ER membrane (Figure 4), and then is apparently rapidly degraded by proteases. Cytochrome P450, an integral ER membrane protein, does not completely cross the membrane. Instead, it resides in the membrane with its signal peptide intact. Its passage through the membrane is prevented by a sequence of amino acids called a halt- or stop-transfer signal. Secretory proteins and proteins destined for membranes distal to the ER completely traverse the membrane bilayer and are discharged into the lumen of the ER. *N*-Glycan chains, if present, are added as these proteins traverse the inner part of the ER membrane—a process called "cotranslational glycosylation." Subsequently, the proteins are found in the lumen of the Golgi apparatus, where further changes in glycan chains occur (Figure 9) prior to intracellular distribution or secretion. There is strong evidence that the signal peptide is involved in the process of protein insertion into ER membranes. Mutant proteins, containing altered signal peptides in which a hydrophobic amino acid is replaced by a hydrophilic one, are not inserted into ER membranes. Nonmembrane proteins (eg, α-globin) to which signal peptides have been attached by genetic engineering can be inserted into the lumen of the ER or even secreted.

There is considerable evidence that a second transposon in the ER membrane is involved in **retrograde transport** of various molecules from the ER lumen to the cytosol. These molecules include unfolded or misfolded glycoproteins, glycopeptides, and oligosaccharides. Some at least of these molecules are degraded in proteasomes. Thus, there is two-way traffic across the ER membrane.

PROTEINS FOLLOW SEVERAL ROUTES TO BE INSERTED INTO OR ATTACHED TO THE MEMBRANES OF THE ENDOPLASMIC RETICULUM

The routes that proteins follow to be inserted into the membranes of the ER include the following.

A. COTRANSLATIONAL INSERTION

Figure 5 shows a variety of ways in which proteins are distributed in the plasma membrane. In particular, the amino terminals of certain proteins (eg, the LDL receptor) can be seen to be on the extracytoplasmic face, whereas for other proteins (eg, the asialoglycoprotein receptor) the carboxyl terminals are on this face. To explain these dispositions, one must consider the initial biosynthetic events at the ER membrane. The **LDL receptor** enters the ER membrane in a manner analogous to a secretory protein (Figure 4); it partly traverses the ER membrane, its signal peptide is cleaved, and its amino terminal protrudes into the lumen. However, it is retained in the membrane because it contains a highly hydrophobic segment, the **halt- or stop-transfer signal.**

This sequence forms the single transmembrane segment of the protein and is its membrane-anchoring domain. The small patch of ER membrane in which the newly synthesized LDL receptor is located subsequently buds off as a component of a transport vesicle, probably from the transitional elements of the ER (Figure 2).

As described below in the discussion of asymmetry of proteins and lipids in membrane assembly, the disposition of the receptor in the ER membrane is preserved in the vesicle, which eventually fuses with the plasma membrane. In contrast, the **asialoglycoprotein receptor** possesses an internal insertion sequence, which inserts into the membrane but is not cleaved. This acts as an anchor, and its carboxyl terminal is extruded through the membrane.

The more complex disposition of the **transporters** (eg, for glucose) can be explained by the fact that alternating transmembrane α-helices act as uncleaved insertion sequences and as halt-transfer signals, respectively. Each pair of helical segments is inserted as a hairpin. Sequences that determine the structure of a protein in a membrane are called **topogenic sequences.** As explained in the legend to Figure 46–5, the above three proteins are examples of **type I, type II,** and **type III** transmembrane proteins.

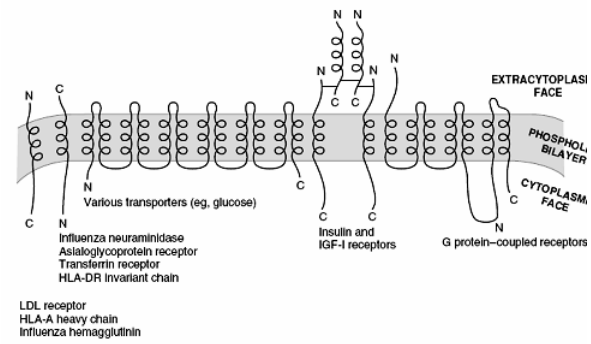


Figure 5. Variations in the way in which proteins are inserted into membranes. This schematic representation, which illustrates a number of possible orientations, shows the segments of the proteins within the membrane as α-helices and the other segments as lines. The LDL receptor, which crosses the membrane once and has its amino terminal on the exterior, is called a type I transmembrane protein. The asialoglycoprotein receptor, which also crosses the membrane once but has its carboxyl terminal on the exterior, is called a type II transmembrane protein. The various transporters indicated (eg, glucose) cross the membrane a number of times and are called type III transmembrane proteins; they are also referred to as polytopic membrane proteins. (N, amino terminal; C, carboxyl terminal.)

B. SYNTHESIS ON FREE POLYRIBOSOMES & SUBSEQUENT ATTACHMENT TO THE ENDOPLASMIC RETICULUMMEMBRANE

An example is cytochrome *b*5, which enters the ER membrane spontaneously.

C. RETENTION AT THE LUMINAL ASPECT OF THE ENDOPLASMIC RETICULUM BY SPECIFIC AMINO ACID SEQUENCES

A number of proteins possess the amino acid sequence **KDEL** (Lys-Asp-Glu-Leu) at their carboxyl terminal. This sequence specifies that such proteins will be attached to the inner aspect of the ER in a relatively loose manner. The chaperone BiP (see below) is one such protein. Actually, KDEL-containing proteins first travel to the Golgi, interact there with a specific KDEL receptor protein, and then return in transport vesicles to the ER, where they dissociate from the receptor.

D. RETROGRADE TRANSPORT FROM THE GOLGI APPARATUS

Certain other non-KDEL-containing proteins destined for the membranes of the ER also pass to the Golgi and then return, by retrograde vesicular transport, to the ER to be inserted therein (see below).

The foregoing paragraphs demonstrate that a **variety of routes** are involved in assembly of the proteins of the ER membranes; a similar situation probably holds for other membranes (eg, the mitochondrial membranes and the plasma membrane). Precise targeting sequences have been identified in some instances (eg, KDEL sequences).

PROTEINS MOVE THROUGH CELLULAR COMPARTMENTS TO SPECIFIC DESTINATIONS

A scheme representing the possible flow of proteins along the ER → Golgi apparatus → plasma membrane route is shown in Figure 6.

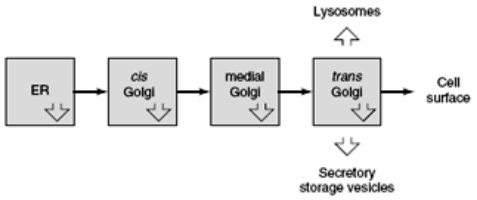


Figure 6. Flow of membrane proteins from the endoplasmic reticulum (ER) to the cell surface. Horizontal arrows denote steps that have been proposed to be signal independent and thus represent bulk flow. The open vertical arrows in the boxes denote retention of proteins that are resident in the membranes of the organelle indicated. The open vertical arrows outside the boxes indicate signal-mediated transport to lysosomes and secretory storage granules.

The horizontal arrows denote transport steps that may be independent of targeting signals, whereas the vertical open arrows represent steps that depend on specific signals. Thus, flow of certain proteins (including membrane proteins) from the ER to the plasma membrane (designated **"bulk flow,"** as it is nonselective) probably occurs without any targeting sequences being involved, ie, by default. On the other hand, insertion of resident proteins into the ER and Golgi membranes is dependent upon specific signals (eg, KDEL or halt-transfer sequences for the ER). Similarly, transport of many enzymes to **lysosomes** is dependent upon the Mannose 6-P signal, and a signal may be involved for entry of proteins into **secretory granules.** Table 4 summarizes information on sequences that are known to be involved in targeting various proteins to their correct intracellular sites.

Table 4. Some sequences or compounds that direct proteins to specific organelles.

Targeting Sequence or Compound	Organelle Targeted
Signal peptide sequence	Membrane of ER
Amino terminal KDEL sequence (Lys-Asp-Glu-Leu)	Luminal surface of ER
Amino terminal sequence (20–80 residues)	Mitochondrial matrix
NLS[1] (eg, Pro_2-Lys_2-Ala-Lys-Val)	Nucleus
PTS[1] (eg, Ser-Lys-Leu)	Peroxisome
Mannose 6-phosphate	Lysosome

[1]NLS, nuclear localization signal; PTS, peroxisomal-matrix targeting sequence.

CHAPERONES ARE PROTEINS THAT PREVENT FAULTY FOLDING & UNPRODUCTIVE INTERACTIONS OF OTHER PROTEINS

Exit from the ER may be the rate-limiting step in the secretory pathway. In this context, it has been found that certain proteins play a role in the assembly or proper folding of other proteins without themselves being components of the latter. Such proteins are called **molecular chaperones;** a number of important properties of these proteins are listed in Table 5, and the names of some of particular importance in the ER are listed in Table 6.

Table 5. Some properties of chaperone proteins.

- Present in a wide range of species from bacteria to humans
- Many are so-called heat shock proteins (Hsp)
- Some are inducible by conditions that cause unfolding of newly synthesized proteins (eg, elevated temperature and various chemicals)
- They bind to predominantly hydrophobic regions of un-folded and aggregated proteins
- They act in part as a quality control or editing mechanism for detecting misfolded or otherwise defective proteins
- Most chaperones show associated ATPase activity, with ATP or ADP being involved in the protein-chaperone interaction
- Found in various cellular compartments such as cytosol, mitochondria, and the lumen of the endoplasmic reticulum

Table 6. Some chaperones and enzymes involved in folding that are located in the rough endoplasmic reticulum.

- BiP (immunoglobulin heavy chain binding protein)
- GRP94 (glucose-regulated protein)
- Calnexin
- Calreticulin
- PDI (protein disulfide isomerase)
- PPI (peptidyl prolyl cis-trans isomerase)

Basically, they stabilize unfolded or partially folded intermediates, allowing them time to fold properly, and event inappropriate interactions, thus combating the formation of nonfunctional structures. Most chaperones exhibit **ATPase activity** and bind ADP and ATP. This activity is important for their effect on folding. The ADP-chaperone complex often has a high affinity for the unfolded protein, which, when bound, stimulates release of ADP with replacement by ATP. The ATP-chaperone complex, in turn, releases segments of the protein that have folded properly, and the cycle involving ADP and ATP binding is repeated until the folded protein is released.

Several examples of chaperones were introduced above when the sorting of mitochondrial proteins was discussed. The **immunoglobulin heavy chain binding protein (BiP)** is located in the lumen of the ER. This protein will bind abnormally folded immunoglobulin heavy chains and certain other proteins and prevent them from leaving the ER, in which they are degraded. Another important chaperone is **calnexin,** a calcium binding protein located in the ER membrane. This protein binds a wide variety of proteins, including mixed histocompatibility (MHC) antigens and a variety of serum proteins. Calnexin binds the mono glycosylated species of glycoproteins that occur during processing of glycoproteins, retaining them in the ER until the glycoprotein has folded properly. **Calreticulin,** which is also a calcium-binding protein, has properties similar to those of calnexin; it is not membrane-bound. Chaperones are not the only proteins in the ER lumen that are concerned with proper folding of proteins. Two enzymes are present that play an active role in folding. **Protein disulfide isomerase (PDI)** promotes rapid reshuffling of disulfide bonds until the correct set is achieved. **Peptidyl prolyl isomerase (PPI)** accelerates folding of proline-containing proteins by catalyzing the cis-trans isomerization of X-Pro bonds, where X is any amino acid residue.

TRANSPORT VESICLES ARE KEY PLAYERS IN INTRACELLULAR PROTEIN TRAFFIC

Most proteins that are synthesized on membrane bound polyribosomes and are destined for the Golgi apparatus or plasma membrane reach these sites inside transport vesicles. The precise mechanisms by which proteins synthesized in the rough ER are inserted into these vesicles are not known. Those involved in transport from the ER to the Golgi apparatus and vice versa—and from the Golgi to the plasma membrane— are mainly clathrin-free, unlike the coated vesicles involved in endocytosis. For the sake of clarity, the non-clathrin-coated vesicles will be referred to in this text as **transport vesicles.** There is evidence that proteins destined for the membranes of the Golgi apparatus contain specific signal sequences. On the other hand, most proteins destined for the plasma membrane or for secretion do not appear to contain specific signals, reaching these destinations by default.

The Golgi Apparatus Is Involved in Glycosylation & Sorting of Proteins

The Golgi apparatus plays two important roles in membrane synthesis. First, it is involved in the **processing of the oligosaccharide chains** of membrane and other N-linked glycoproteins and also contains enzymes involved in O-glycosylation. Second, it is involved in the **sorting** of various proteins prior to their delivery to their appropriate intracellular destinations.

All parts of the Golgi apparatus participate in the first role, whereas the trans-Golgi is particularly involved in the second and is very rich in vesicles. Because of their central role in protein transport, considerable research has been conducted in recent years concerning the formation and fate of transport vesicles.

A Model of Non-Clathrin-Coated Vesicles Involves SNAREs & Other Factors

Vesicles lie at the heart of intracellular transport of many proteins. Recently, significant progress has been made in understanding the events involved in vesicle formation and transport. This has transpired because of the use of a number of approaches. These include establishment of **cell-free systems** with which to study vesicle formation. For instance, it is possible to observe, by electron microscopy, budding of vesicles from Golgi preparations incubated with cytosol and ATP. The development of genetic approaches for studying vesicles in yeast has also been crucial. The picture is complex, with its own nomenclature (Table 7), and involves a variety of cytosolic and membrane proteins, GTP, ATP, and accessory factors.

Table 7. Factors involved in the formation of non-clathrin-coated vesicles and their transport.

- ARF: ADP-ribosylation factor, a GTPase
- Coatomer: A family of at least seven coat proteins (α, β, γ, δ, ε, β′, and ζ). Different transport vesicles have different complements of coat proteins.
- SNAP: Soluble NSF attachment factor
- SNARE: SNAP receptor
- v-SNARE: Vesicle SNARE
- t-SNARE: Target SNARE
- GTP-γ-S: A nonhydrolyzable analog of GTP, used to test the involvement of GTP
- NEM: *N*-Ethylmaleimide, a chemical that alkylates sulfhydryl groups
- NSF: NEM-sensitive factor, an ATPase
- Rab proteins: A family of ras-related proteins first observed in rat brain; they are GTPases and are active when GTP is found
- Sec1: A member of a family of proteins that attach to t-SNAREs and are displaced from them by Rab proteins, thereby allowing v-SNARE–t-SNARE interactions to occur.

Based largely on a proposal by Rothman and colleagues, anterograde vesicular transport can be considered to occur in eight steps (Figure 7). The basic concept is that each transport vesicle bears a unique address marker consisting of one or more **v-SNARE proteins,** while each target membrane bears one or more **complementary t-SNARE proteins** with which the former interact specifically.

Step 1: **Coat assembly** is initiated when ARF is activated by binding GTP, which is exchanged for GDP. This leads to the association of GTP-bound ARF with its putative receptor (hatched in Figure 7) in the donor membrane.
Step 2: Membrane-associated ARF **recruits the coat proteins** that comprise the coatomer shell from the cytosol, forming a coated bud.
Step 3: The **bud pinches off** in a process involving acyl-CoA—and probably ATP—to complete the formation of the coated vesicle.
Step 4: **Coat disassembly** (involving dissociation of ARF and coatomer shell) follows hydrolysis of bound GTP; uncoating is necessary for fusion to occur.
Step 5: **Vesicle targeting** is achieved via members of a family of integral proteins, termed v-SNAREs, that tag the vesicle during its budding. v-SNAREs pair with cognate t-SNAREs in the target membrane to dock the vesicle. It is presumed that steps 4 and 5 are closely coupled and that step 4 may follow step 5, with ARF and the coatomer shell rapidly dissociating after docking.
Step 6: The **general fusion machinery** then assembles on the paired SNARE complex; it includes an ATPase (NSF; NEM-sensitive factor) and the SNAP (soluble NSF attachment factor) proteins. SNAPs bind to the SNARE (SNAP receptor) complex, enabling NSF to bind.
Step 7: **Hydrolysis of ATP by NSF** is essential for fusion, a process that can be inhibited by NEM (*N*ethylmaleimide). Certain other proteins and calcium are also required.
Step 8: **Retrograde transport** occurs to restart the cycle. This last step may retrieve certain proteins or recycle v-SNAREs. Nocodazole, a microtubule disrupting agent, inhibits this step.

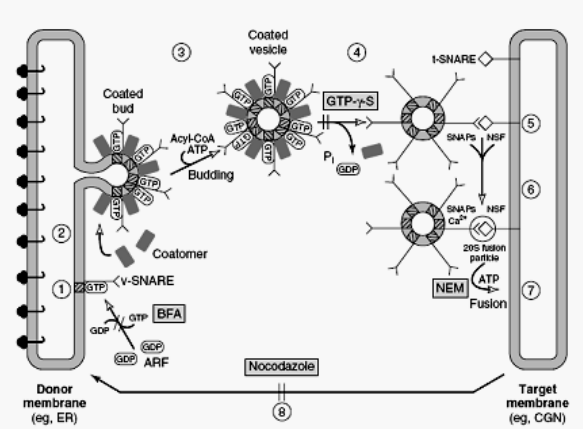


Figure 7. Model of the steps in a round of anterograde vesicular transport. The cycle starts in the bottom left-hand side of the figure, where two molecules of ARF are represented as small ovals containing GDP. The steps in the cycle are described in the text. Most of the abbreviations used are explained in Table 7. . The roles of Rab and Sec1 proteins (see text) in the overall process are not dealt with in this figure. (CGN, cis-Golgi network; BFA, Brefeldin A.)

Brefeldin A Inhibits the Coating Process

The following points expand and clarify the above.

(a) To participate in step 1, ARF must first be modified by addition of **myristic acid** (C14:0), employing myristoyl-CoA as the acyl donor. Myristoylation is one of a number of enzyme-catalyzed posttranslational modifications, involving addition of certain lipids to specific residues of proteins, that facilitate the binding of proteins to the cytosolic surfaces of membranes or vesicles. Others are addition of palmitate, farnesyl, and geranylgeranyl; the two latter molecules are polyisoprenoids containing 15 and 20 carbon atoms, respectively.
(b) At least three different types of coated vesicles have been distinguished: **COPI, COPII,** and **clathrincoated vesicles;** the first two are referred to here as transport vesicles. Many other types of vesicles no doubt remain to be discovered. COPI vesicles are involved in bidirectional transport from the ER to the Golgi and in the reverse direction, whereas COPII vesicles are involved mainly in transport in the former direction. Clathrin-containing vesicles are involved in transport from the trans-Golgi network to prelysosomes and from the plasma membrane to endosomes, respectively. Regarding **selection** of cargo molecules by vesicles, this appears to be primarily a function of the coat proteins of vesicles. **Cargo molecules** may interact with coat proteins either directly or via intermediary proteins that attach to coat proteins, and they then become enclosed in their appropriate vesicles.
(c) The fungal metabolite **brefeldin A** prevents GTP from binding to ARF in step 1 and thus inhibits the entire coating process. In its presence, the Golgi apparatus appears to disintegrate, and fragments are lost. It may do this by inhibiting the guanine nucleotide exchanger involved in step 1.
(d) GTP-γ-S (a nonhydrolyzable analog of GTP often used in investigations of the role of GTP in biochemical processes) blocks disassembly of the coat from coated vesicles, leading to a build-up of coated vesicles.

(e) A family of Ras-like proteins, called the **Rab protein family,** are required in several steps of intracellular protein transport, regulated secretion, and endocytosis. They are small monomeric GTPases that attach to the cytosolic faces of membranes via geranylgeranyl chains. They attach in the GTP-bound state (not shown in Figure 7) to the budding vesicle. Another family of proteins **(Sec1)** binds to t-SNAREs and prevents interaction with them and their complementary v-SNAREs. When a vesicle interacts with its target membrane, Rab proteins displace Sec1 proteins and the v-SNAREt- SNARE interaction is free to occur. It appears that the Rab and Sec1 families of proteins regulate the speed of vesicle formation, opposing each other. Rab proteins have been likened to throttles and Sec1 proteins to dampers on the overall process of vesicle formation.

(f) Studies using v- and t-SNARE proteins reconstituted into separate lipid bilayer vesicles have indicated that they form **SNAREpins,** ie, SNARE complexes that link two membranes (vesicles). SNAPs and NSF are required for formation of SNAREpins, but once they have formed they can apparently lead to spontaneous fusion of membranes at physiologic temperature, suggesting that they are the minimal machinery required for membrane fusion.

(g) The fusion of synaptic vesicles with the plasma membrane of **neurons** involves a series of events similar to that described above. For example, one v-SNARE is designated **synaptobrevin** and two t-SNAREs are designated **syntaxin** and **SNAP 25** (synaptosome-associated protein of 25 kDa). **Botulinum B toxin** is one of the most lethal toxins known and the most serious cause of food poisoning. One component of this toxin is a protease that appears to cleave only synaptobrevin, thus inhibiting release of acetylcholine at the neuromuscular junction and possibly proving fatal, depending on the dose taken.

(h) Although the above model describes nonclathrin- coated vesicles, it appears likely that many of the events described above apply, at least in principle, to clathrin-coated vesicles.

THE ASSEMBLY OF MEMBRANES IS COMPLEX

There are many cellular membranes, each with its own specific features. No satisfactory scheme describing the assembly of any one of these membranes is available. How various proteins are initially inserted into the membrane of the ER has been discussed above. The transport of proteins, including membrane proteins, to various parts of the cell inside vesicles has also been described. Some general points about membrane assembly remain to be addressed.

Asymmetry of Both Proteins & Lipids Is Maintained During Membrane Assembly

Vesicles formed from membranes of the ER and Golgi apparatus, either naturally or pinched off by homogenization, exhibit **transverse asymmetries** of both lipid and protein. These asymmetries are maintained during fusion of transport vesicles with the plasma membrane. The inside of the vesicles after fusion becomes the outside of the plasma membrane, and the cytoplasmic side of the vesicles remains the cytoplasmic side of the membrane (Figure 8). Since the transverse asymmetry of the membranes already exists in the vesicles of the ER well before they are fused to the plasma membrane, a major problem of membrane assembly becomes understanding how the integral proteins are inserted into the lipid bilayer of the ER.

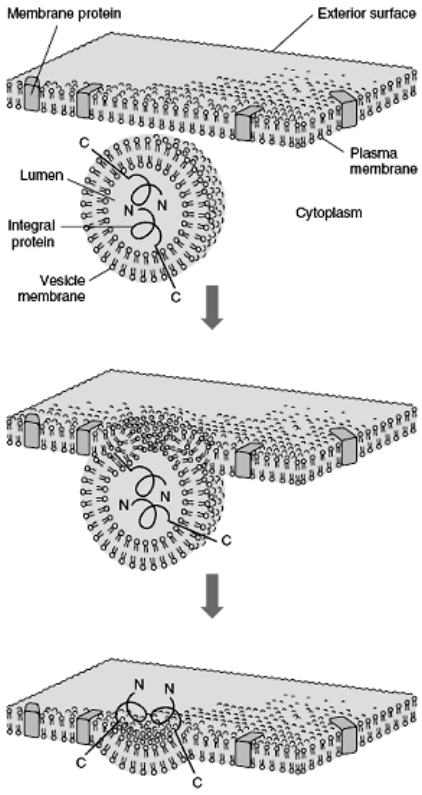


Figure 8. Fusion of a vesicle with the plasma membrane preserves the orientation of any integral proteins embedded in the vesicle bilayer. Initially, the amino terminal of the protein faces the lumen, or inner cavity, of such a vesicle. After fusion, the amino terminal is on the exterior surface of the plasma membrane. That the orientation of the protein has not been reversed can be perceived by noting that the other end of the molecule, the carboxyl terminal, is always immersed in the cytoplasm. The lumen of a vesicle and the outside of the cell are topologically equivalent.

Phospholipids are the major class of lipid in membranes. The enzymes responsible for the synthesis of phospholipids reside in the cytoplasmic surface of the cisternae of the ER. As phospholipids are synthesized at that site, they probably self-assemble into thermodynamically stable bimolecular layers, thereby expanding the membrane and perhaps promoting the detachment of so-called lipid vesicles from it. It has been proposed that these vesicles travel to other sites, donating their lipids to other membranes; however, little is known about this matter. As indicated above, cytosolic proteins that take up phospholipids from one membrane and release them to another (ie, phospholipid exchange proteins) have been demonstrated; they probably play a role in contributing to the specific lipid composition of various membranes.

Lipids & Proteins Undergo Turnover at Different Rates in Different Membranes

It has been shown that the half-lives of the lipids of the ER membranes of rat liver are generally shorter than those of its proteins, so that the **turnover rates of lipids and proteins are independent.** Indeed, different lipids have been found to have different half-lives. Furthermore, the half-lives of the proteins of these membranes vary quite widely, some exhibiting short (hours) and others long (days) half-lives. Thus, individual lipids and proteins of the ER membranes appear to be inserted into it relatively independently; this is the case for many other membranes.

The biogenesis of membranes is thus a complex process about which much remains to be learned. One indication of the complexity involved is to consider the number of **posttranslational modifications** that membrane proteins may be subjected to prior to attaining their mature state. These include proteolysis, assembly into multimers, glycosylation, addition of a glycophosphatidylinositol (GPI) anchor, sulfation on tyrosine or carbohydrate moieties, phosphorylation, acylation, and prenylation—a list that is undoubtedly not complete. Nevertheless, significant progress has been made; Table 8 summarizes some of the major features of membrane assembly that have emerged to date.

A5. Electrical properties of membranes

Biological membranes serve as barriers to the passage of ions and polar molecules, a fact that is reflected in their high electrical resistance and capacitance. The electrical resistance is usually 10^3 ohms cm^{-2}, while the capacitance is 0.5 – 1.5 microfarad (μF) cm^{-2}. The corresponding values for artificial membranes are ~ 10^7 ohms cm^{-2} and 0.6 – 0.9 μF cm^{-2}. The lower resistance of biological membranes must result from the presence of proteins and other ion-carrying substances or of pores in the membranes. The capacitance values for the two types of membrane are very close to those expected for a bilayer with a thickness of ~ 2.5 nm and a dielectric constant of 2. The electrical potential gradient is steep.

Outer cell surfaces usually carry a net negative charge, the result of phosphate groups of phospholipids, of carboxylate groups on proteins, and of sialic acids attached to glycoproteins. This negatively charged surface layer attracts ions of the opposite charge (counterions), including protons, and repels those of the same charge. The result is development of a diffuse **electrical double layer** consisting of the fixed negative charges on the surface and a positive **ionic atmosphere** extending into the solution for a distance that depends upon the ionic strength. This ionic atmosphere is analogous to that postulated by the Debye–Hückel theory. At the physiological ionic strength of 0.145 M the thickness of the double layer, taken as the distance at which the electrical potential falls to a certain fraction of that at the cell surface, is about 0.8 nm. However, the double layer thickness increases to about three times this value at an ionic strength of 10^{-3} M and to still greater distances at lower ionic strengths.

The net surface charge of a cell and the associated electrical double layer are important in interactions between cells and may influence the development of extracellular structure such as basement membranes. The net negative charge on cells also gives rise to an experimentally measurable electrophoretic mobility. A characteristic of living cells is the maintenance of **ionic gradients** across the plasma membrane. Thus almost all cells accumulate K^+, even "pumping" it from very dilute external solutions. Cells also exclude sodium, pumping it out from the cytoplasm. If a microelectrode is inserted through a cell membrane and the potential difference is measured between the inside and outside of the cell, a **resting potential** which, in nerve cells, may be as high as 90 mV is observed. The origin of the potential appears to lie in the concentration differences of ions. From the value of ΔG for dilution of an ion (Eq. 1) and the relationship between ΔG and electrode potential (Eq. 2), the Nernst equation (Eq. 3) can be derived.

$$\Delta \bar{G} \text{ (dilution from } a_1 \text{ to } a_2) = RT \ln (a_2/a_1) \quad \textbf{(1)}$$

When electrons flow in the external circuit the maximum amount of work that can be done per mole of electrochemical reaction ($-\Delta G$) is given by Eq. 2

$$\Delta G = nEF = nE \times 96.487 \text{ kJ mol}^{-1} \text{ V}^{-1} = -nE \times 23.061 \text{ kcal mol}^{-1} \text{ V}^{-1} \quad \textbf{(2)}$$

$$E_m = \frac{RT}{nF} \ln \left(\frac{c_1}{c_2}\right) = \frac{0.059}{n} \log \left(\frac{c_1}{c_2}\right) \quad \text{at } 25°C \quad \textbf{(3)}$$

According to this equation, which applies to a single ion for which the membrane is permeable, a 10-fold concentration difference across the membrane for a monovalent ion (n = 1) would lead to a 59-mVmembrane potential, Em. Since membranes are relatively impermeable to sodium ions, it is generally conceded that for many membranes the origin of the membrane potential lies mainly with the potassium ion concentration difference which is maintained by the Na^+, K^+-ATPase. A more complete equation takes account of K^+, Na^+, and Cl^- together with their respective permeabilities. Note also that Eq. 4 is also often called the Nernst equation. Protons are also pumped across cytoplasmic and inner mitochondrial membranes.

$$E = E^\circ + \frac{RT}{nF} \ln \frac{[A][H^+]^2}{[AH_2]} \quad \textbf{(4)}$$

If $n = 2$

$$E = E^\circ + 0.0296 \log \frac{[A][H^+]^2}{[AH_2]} \quad \text{volts at } 25°C$$

The flow of protons from inside to outside also contributes to the membrane potential. The positive charges of H^+, K^+, and other cations associated with the external membrane surface are balanced by the negative charges of protein molecules as well as Cl^- and phosphate anions that are in or near to the inner surface of the membranes. Another possibility for proton flow has intrigued biophysicists for years. Membranes often display a substantial electrical conductivity in a lateral direction along the membrane surface. Electrical conduction may involve movement of protons along hydrogen bonded lines, e.g., involving ethanolamine head groups or phosphate groups and bridging water as previously discussed. Alternatively, conduction may depend upon membrane-associated proteins. This lateral proton conduction may be important to many proton-driven membrane processess, such as rotation of bacterial flagella, ATP synthesis, and pumping of ions.

B. Structural organization and function of intracellular organelles:

Cells are the structural and functional units of life. The smallest organisms are composed of only a single cell while the largest are made up of billions of cells. Even when comparing the most diverse and complex organisms, at the cellular level they are remarkably similar. Even though the human body has over 100 different cell types, they all share certain features and they even have many characteristics in common with plants.

From Procaryotes to Eucaryotes: It is thought that all organisms living now on earth derive from a single primordial cell born more than 3 billion years ago. This cell, out-reproducing its competitors, took the lead in the process of cell division and evolution that eventually covered the earth with green, changed the composition of its atmosphere, and made it the home of intelligent life. The family resemblances among all organisms seem too strong to be explained in any other way. One important landmark along this evolutionary road occurred about 1.5 billion years ago, when there was a transition from small cells with relatively simple internal structures - the so-called procaryotic cells, which include the various types of bacteria - to a flourishing of larger and radically more complex eucaryotic cells such as are found in higher animals and plants.

Procaryotic Cells Are Structurally Simple but Biochemically Diverse

Bacteria are the simplest organisms found in most natural environments. They are spherical or rod-shaped cells, commonly several micrometers in linear dimension. They often possess a tough protective coat, called a cell wall, beneath which a plasma membrane encloses a single cytoplasmic compartment containing DNA, RNA, proteins, and small molecules. In the electron microscope this cell interior appears as a matrix of varying texture without any obvious organized internal structure.

Bacteria are small and can replicate quickly, simply dividing in two by binary fission. When food is plentiful, "survival of the fittest" generally means survival of those that can divide the fastest. Under optimal conditions a single procaryotic cell can divide every 20 minutes and thereby give rise to 5 billion cells (approximately equal to the present human population on earth) in less than 11 hours. The ability to divide quickly enables populations of bacteria to adapt rapidly to changes in their environment. Under laboratory conditions, for example, a population of bacteria maintained in a large vat will evolve within a few weeks by spontaneous mutation and natural selection to utilize new types of sugar molecules as carbon sources.

In nature bacteria live in an enormous variety of ecological niches, and they show a corresponding richness in their underlying biochemical composition. Two distantly related groups can be recognized: the eubacteria, which are the commonly encountered forms that inhabit soil, water, and larger living organisms; and the archaebacteria, which are found in such incommodious environments as bogs, ocean depths, salt brines, and hot acid springs.

There are species of bacteria that can utilize virtually any type of organic molecule as food, including sugars, amino acids, fats, hydrocarbons, polypeptides, and polysaccharides. Some are even able to obtain their carbon atoms from CO_2 and their nitrogen atoms from N_2. Despite their relative simplicity, bacteria have existed for longer than any other organisms and still are the most abundant type of cell on earth.

Eucaryotic Cells Contain Several Distinctive Organelles

As molecular oxygen accumulated in the atmosphere, what happened to the remaining anaerobic organisms with which life had begun? In a world that was rich in oxygen, which they could not use, they were at a severe disadvantage. Some, no doubt, became extinct. Others either developed a capacity for respiration or found niches in which oxygen was largely absent, where they could continue an anaerobic way of life. Others became predators or parasites on aerobic cells. And some, it seems, hit upon a strategy for survival more cunning and vastly richer in implications for the future: they are believed to have formed an intimate association with an aerobic type of cell, living with it in symbiosis. This is the most plausible explanation for the metabolic organization of present-day cells of the eucaryotic type with which this book will be chiefly concerned.

Eucaryotic cells, by definition and in contrast to procaryotic cells, have a nucleus (caryon in Greek), which contains most of the cell's DNA, enclosed by a double layer of membrane. The DNA is thereby kept in a compartment separate from the rest of the contents of the cell, the cytoplasm, where most of the cell's metabolic reactions occur. In the cytoplasm, moreover, many distinctive organelles can be recognized. Prominent among these are two types of small bodies, the chloroplasts and mitochondria. Each of these is enclosed in its own double layer of membrane, which is chemically different from the membranes surrounding the nucleus. Mitochondria are an almost universal feature of eucaryotic cells, whereas chloroplasts are found only in those eucaryotic cells that are capable of photosynthesis - that is, in plants but not in animals or fungi. Both organelles almost certainly have a symbiotic origin.

B1. Cell wall

A cell wall is a fairly rigid layer surrounding a cell, located external to the cell membrane, which provides the cell with structural support, protection, and acts as a filtering mechanism. The cell wall also prevents over-expansion when water enters the cell. They are found in plants, bacteria, fungi, algae, and some archaea. Animals, and protozoa do not have cell walls.

The materials in a cell wall varies between species. In plants, the strongest component of the complex cell wall is a carbohydrate polymer called cellulose. In bacteria, peptidoglycan forms the cell wall. Archaean cell walls have various compositions, and may be formed of glycoprotein S-layers, pseudopeptidoglycan, or polysaccharides. Fungi possess cell walls of chitin, and algae typically possess walls constructed of glycoproteins and polysaccharides, however certain algal species may have a cell wall composed of silicic acid. Often, other accessory molecules are found anchored to the cell wall.

Properties

The cell wall serves a similar purpose in those organisms that possess them. The wall gives cells rigidity and strength, offering protection against mechanical stress. In multicellular organisms, it permits the organism to build and hold its shape (morphogenesis). The cell wall also limits the entry of large molecules that may be toxic to the cell. It further permits the creation of a stable osmotic environment by preventing osmotic lysis and helping to retain water. The composition, properties, and form of the cell wall may change during the cell cycle and depend on growth conditions.

Rigidity

In most cells, the cell wall is semi-rigid, meaning that it will bend somewhat rather than holding a fixed shape. This flexibility is seen when plants wilt, so that the stems and leaves begin to droop, or in seaweeds that bend in water currents. Wall rigidity seen in healthy plants results from a combination of the wall construction and turgor pressure. The rigidity of the cell wall thus results in part from inflation of the cell contained. This inflation is a result of the passive uptake of water.

Other cell walls are inflexible. In plants, a secondary cell wall is a thicker additional layer of cellulose. Additional layers may be formed containing lignin in xylem cell walls, or containing suberin in cork cell walls. These compounds are rigid and waterproof, making the secondary wall stiff. Both wood and bark cells of trees have secondary walls. Other parts of plants such as the leaf stalk may acquire similar reinforcement to resist the strain of physical forces.

Certain single-cell protists and algae also produce a rigid wall. Diatoms build a frustule from silica extracted from the surrounding water; radiolarians also produce a test from minerals. Many green algae, such as the Dasycladales encase their cells in a secreted skeleton of calcium carbonate. In each case, the wall is rigid and essentially inorganic.

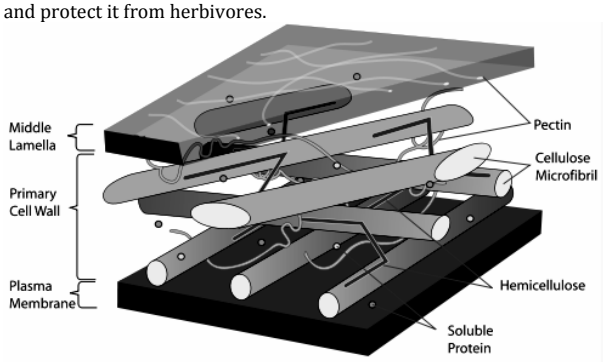
Permeability

The primary cell wall of most plant cells is semi-permeable and permit the passage of small molecules and small proteins, with size exclusion estimated to be 30-60 kDa. Key nutrients, especially water and carbon dioxide, are distributed throughout the plant from cell wall to cell wall in apoplastic flow.

I. Plant cell walls

Composition: The major carbohydrates making up the primary (growing) cell wall are cellulose, hemicellulose and pectin. The cellulose microfibrils are linked via hemicellulosic tethers to form the cellulose-hemicellulose network, which is embedded in the pectin matrix. The most common hemicellulose in the primary cell wall is xyloglucan. In grass cell walls, xyloglucan and pectin are reduced in abundance and partially replaced by glucuronarabinoxylan, a hemicellulose. Primary cell walls characteristically extend (grow) by a mechanism called acid growth, which involves turgor-driven movement of the strong cellulose microfibrils within the weaker hemicellulose/pectin matrix, catalyzed by expansin proteins.

The major polymers that make up wood (largely secondary cell walls) include cellulose (35 to 50%), xylan, a type of hemicellulose, (20 to 35%) and a complex phenolic polymer called lignin (10 to 25%). Lignin penetrates the spaces in the cell wall between cellulose, hemicellulose and pectin components, driving out water and strengthening the wall. Secondary walls - especially in grasses - may also contain microscopic silicate crystals, which may strengthen the wall and protect it from herbivores.



Plant cells walls also contain numerous enzymes, such as hydrolases, esterases, peroxidases, and transglycosylases, that cut, trim and cross link wall polymers. Small amounts (1-5%) of structural proteins are found in most plant cell walls; they are classified as hydroxyproline-rich glycoproteins (HRGP), arabinogalactan proteins (AGP), glycine-rich proteins (GRPs), and proline-rich proteins (PRPs). Each class of glycoprotein is defined by a characteristic, highly repetitive protein sequence. Most are glycosylated, contain hydroxyproline (Hyp) and become cross-linked in the cell wall. These proteins are often concentrated in specialized cells and in cell corners. Cell walls of the epidermis and endodermis may also contain suberin or cutin, two polyester-like polymers that protect the cell from herbivores. The relative composition of carbohydrates, secondary compounds and protein varies between plants and between the cell type and age.

Up to three strata or layers may be found in plant cell walls:

- The middle lamella, a layer rich in pectins. This outermost layer forming the interface between adjacent plant cells and glues them together.
- The primary cell wall, generally a thin, flexible and extensible layer formed while the cell is growing.
- The secondary cell wall, a thick layer formed inside the primary cell wall after the cell is fully grown. It is not found in all cell types.

In some cells, such as found xylem, the secondary wall contains lignin, which strengthens and waterpoofs the wall.
Cell walls in some plant tissues also function as storage depots for carbohydrates that can be broken down and resorbed to supply the metabolic and growth needs of the plant. For example, endosperm cell walls in the seeds of cereal grasses, nasturtium, and other species, are rich in glucans and other polysaccharides that are readily digested by enzymes during seed germination to form simple sugars that nourish the growing embryo. Cellulose microfibrils are not readily digested by plants, however.

Formation: The middle lamella is laid down first, formed from the cell plate during cytokinesis, and the primary cell wall is then deposited inside the middle lamella. The actual structure of the cell wall is not clearly defined and several models exist - the covalently linked cross model, the tether model, the diffuse layer model and the stratified layer model. However, the primary cell wall, can be defined as composed of cellulose microfibrils aligned at all angles. Microfibrils are held together by hydrogen bonds to provide a high tensile strength. The cells are held together and share the gelatinous membrane called the middle lamella, which contains magnesium and calcium pectates (salts of pectic acid). Cells interact though plasmodesma(ta), which are inter-connecting channels of cytoplasm that connect to the protoplasts of adjacent cells across the cell wall.

In some plants and cell types, after a maximum size or point in development has been reached, a secondary wall is constructed between the plant cell and primary wall. Unlike the primary wall, the microfibrils are aligned mostly in the same direction, and with each additional layer the orientation changes slightly. Cells with secondary cell walls are rigid. Cell to cell communication is possible through pits in the secondary cell wall that allow plasmodesma to connect cells through the secondary cell walls.

Trees modify cell walls in their branches to reinforce and support structure. Conifers, such as pine, produce thicker cell walls on the undersides of branches to push their branches upwards. The resulting wood is called compression wood. By contrast, hardwood trees reinforce the walls on the upper sides of branches to pull their branches up. This is known as tension wood. Additional thickening may occur in other parts of the plant in response to mechanical stress.

II. Algal cell walls

Like plants, algae have cell walls. Algal cell walls contain cellulose and a variety of glycoproteins. The inclusion of additional polysaccharides in algal cells walls is used as a feature for algal taxonomy.

- Manosyl form microfibrils in the cell walls of a number of marine green algae including those from the genera, Codium, Dasycladus, and Acetabularia as well as in the walls of some red algae, like Porphyra and Bangia.
- Xylanes
- Alginic acid is a common polysaccharide in the cell walls of brown algae

- Sulfonated polysaccharides occur in the cell walls of most algae; those common in red algae include agarose, carrageenan, porphyran, furcelleran and funoran.
- Other compounds that may accumulate in algal cell walls include sporopollenin and calcium ions.

The group of algae known as the diatoms synthesize their cell walls (also known as frustules or valves) from silicic acid (specifically orthosilicic acid, H_4SiO_4). The acid is polymerised intra-cellularly, then the wall is extruded to protect the cell. Significantly, relative to the organic cell walls produced by other groups, silica frustules require less energy to synthesize (approximately 8%), potentially a major saving on the overall cell energy budget and possibly an explanation for higher growth rates in diatoms.

III. Fungal cell walls

There are several groups of organisms that may be called "fungi". Some of these groups have been transferred out of the Kingdom Fungi, in part because of fundamental biochemical differences in the composition of the cell wall. Most true fungi have a cell wall consisting largely of chitin and other polysaccharides.[8] True fungi do not have cellulose in their cell walls, but some fungus-like organisms do.

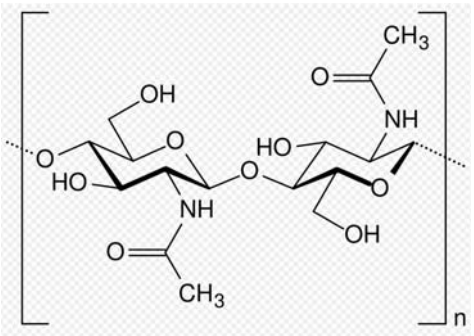


Fig 2: Chemical structure of a unit from a chitin polymer chain.

True fungi: Not all species of fungi have cell walls but in those that do, the plasma membrane is followed by three layers of cell wall material. From inside out these are:

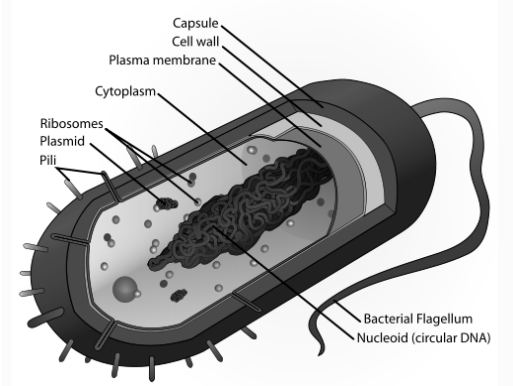
- a chitin layer (polymer consisting mainly of unbranched chains of N-acetyl-D-glucosamine)
- a layer of β-1,3-glucan
- a layer of mannoproteins (mannose-containing glycoproteins) which are heavily glycosylated at the outside of the cell.

Fungus-like protists: The group Oomycetes, also known as water molds, are saprotrophic plant pathogens like fungi. Until recently they were widely believed to be fungi, but structural and molecular evidence has led to their reclassification as heterokonts, related to autotrophic brown algae and diatoms. Unlike fungi, oomycetes typically possess cell walls of cellulose and glucans rather than chitin, although some genera (such as Achlya and Saprolegnia) do have chitin in their walls. The fraction of cellulose in the walls is no more than 4 to 20%, far less than the fraction comprised by glucans. Oomycete cell walls also contain the amino acid hydroxyproline, which is not found in fungal cell walls.

The dictyostelids are another group formerly classified among the fungi. They are slime moulds that feed as unicellular amoebae, but aggregate into a reproductive stalk and sporangium under certain conditions. Cells of the reproductive stalk, as well as the spores formed at the apex, possess a cellulose wall. The spore wall has been shown to possess three layers, the middle of which is composed primarily of cellulose, and the innermost is sensitive to cellulase and pronase.

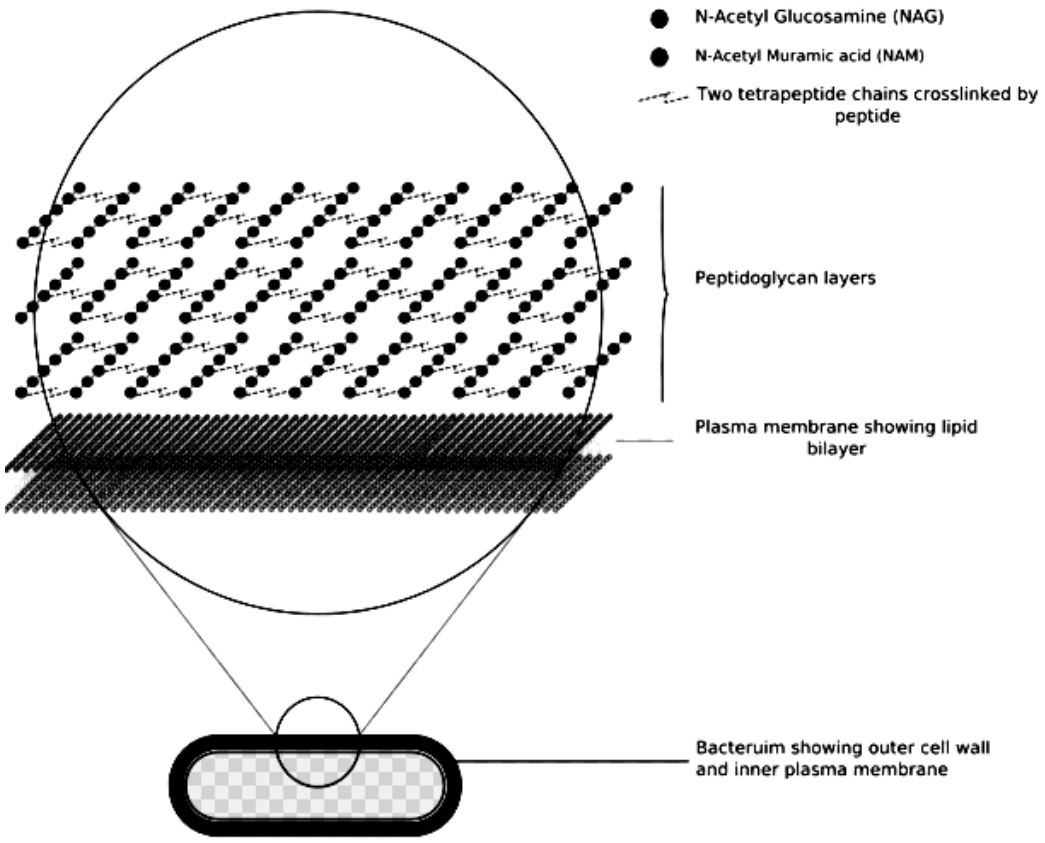
IV. Prokaryotic cell walls

Bacterial cell walls: Around the outside of the cell membrane is the bacterial cell wall. Bacterial cell walls are made of peptidoglycan (also called murein), which is made from polysaccharide chains cross-linked by unusual peptides containing D-amino acids. Bacterial cell walls are different from the cell walls of plants and fungi which are made of cellulose and chitin, respectively. The cell wall of bacteria is also distinct from that of Archaea, which do not contain peptidoglycan. The cell wall is essential to the survival of many bacteria. The antibiotic penicillin is able to kill bacteria by inhibiting a step in the synthesis of peptidoglycan.



There are broadly speaking two different types of cell wall in bacteria, called Gram-positive and Gram-negative. The names originate from the reaction of cells to the Gram stain, a test long-employed for the classification of bacterial species.

Gram-positive bacteria possess a thick cell wall containing many layers of peptidoglycan and teichoic acids. In contrast, Gram-negative bacteria have a relatively thin cell wall consisting of a few layers of peptidoglycan surrounded by a second lipid membrane containing lipopolysaccharides and lipoproteins. Most bacteria have the Gram-negative cell wall and only the Firmicutes and Actinobacteria (previously known as the low G+C and high G+C Gram-positive bacteria, respectively) have the alternative Gram-positive arrangement. These differences in structure can produce differences in antibiotic susceptibility, for instance vancomycin can kill only Gram-positive bacteria and is ineffective against Gram-negative pathogens, such as Haemophilus influenzae or Pseudomonas aeruginosa.



Simplified schematic of cell wall in a gram-positive bacterium
(showing plasma membrane; teichoic acids not shown)

Archaeal cell walls: Although not truly unique, the cell walls of Archaea are unusual. Whereas peptidoglycan is a standard component of all bacterial cell walls, all archaeal cell walls lack peptidoglycan, with the exception of one group of methanogens. In that group, the peptidoglycan is a modified form very different from the kind found in bacteria. There are four types of cell wall currently known among the Archaea.

One type of archaeal cell wall is that composed of pseudopeptidoglycan (also called pseudomurein). This type of wall is found in some methanogens, such as Methanobacterium and Methanothermus. While the overall structure of archaeal pseudopeptidoglycan superficially resembles that of bacterial peptidoglycan, there are a number of significant chemical differences. Like the peptidoglycan found in bacterial cell walls, pseudopeptidoglycan consists of polymer chains of glycan cross-linked by short peptide connections. However, unlike peptidoglycan, the sugar N-acetylmuramic acid is replaced by N-acetyltalosaminuronic acid, and the two sugars are bonded with a β,1-3 glycosidic linkage instead of β,1-4. Additionally, the cross-linking peptides are L-amino acids rather than D-amino acids as they are in bacteria.
A second type of archaeal cell wall is found in Methanosarcina and Halococcus. This type of cell wall is composed entirely of a thick layer of polysaccharides, which may be sulfated in the case of Halococcus. Structure in this type of wall is complex and as yet is not fully investigated.

A third type of wall among the Archaea consists of glycoprotein, and occurs in the hyperthermophiles, Halobacterium, and some methanogens. In Halobacterium, the proteins in the wall have a high content of acidic amino acids, giving the wall an overall negative charge. The result is an unstable structure that is stabilized by the presence of large quantities of positive sodium ions that neutralize the charge. Consequently, Halobacterium thrives only under conditions with high salinity.

In other Archaea, such as Methanomicrobium and Desulfurococcus, the wall may be composed only of surface-layer proteins, known as an S-layer. S-layers are common in bacteria, where they serve as either the sole cell-wall component or an outer layer in conjunction with peptidoglycan and murein. Most Archaea are Gram-negative, though at least one Gram-positive member is known.

B2. The Nucleus

The presence of a nucleus is the principal feature that distinguishes eukaryotic from prokaryotic cells. By housing the cell's genome, the nucleus serves both as the repository of genetic information and as the cell's control center. DNA replication, transcription, and RNA processing all take place within the nucleus, with only the final stage of gene expression (translation) localized to the cytoplasm.

By separating the genome from the cytoplasm, the nuclear envelope allows gene expression to be regulated by mechanisms that are unique to eukaryotes. Whereas prokaryotic mRNAs are translated while their transcription is still in process, eukaryotic mRNAs undergo posttranscriptional processing (e.g., splicing) before being transported from the nucleus to the cytoplasm. The

presence of a nucleus thus allows gene expression to be regulated by posttranscriptional mechanisms, such as alternative splicing. By limiting the access of proteins to the genetic material, the nuclear envelope also provides novel opportunities for the control of gene expression at the level of transcription. For example, the expression of some eukaryotic genes is controlled by the regulated transport of transcription factors from the cytoplasm into the nucleus—a form of transcriptional regulation unavailable to prokaryotes. The separation of the genome from the site of mRNA translation thus plays a central role in eukaryotic gene expression.

The Nuclear Envelope and Traffic between the Nucleus and Cytoplasm

The nuclear envelope separates the contents of the nucleus from the cytoplasm and provides the structural framework of the nucleus. The nuclear membranes, acting as barriers that prevent the free passage of molecules between the nucleus and the cytoplasm, maintain the nucleus as a distinct biochemical compartment. The sole channels through the nuclear envelope are provided by the nuclear pore complexes, which allow the regulated exchange of molecules between the nucleus and cytoplasm. The selective traffic of proteins and RNAs through the nuclear pore complexes not only establishes the internal composition of the nucleus, but also plays a critical role in regulating eukaryotic gene expression.

Structure of the Nuclear Envelope

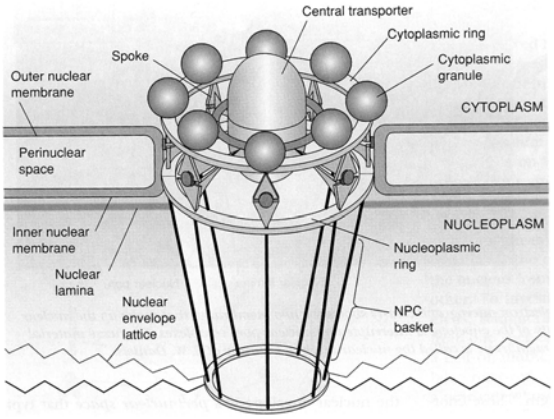
The nuclear envelope has a complex structure, consisting of two nuclear membranes, an underlying nuclear lamina, and nuclear pore complexes (Figure). The nucleus is surrounded by a system of two concentric membranes, called the inner and outer nuclear membranes. The outer nuclear membrane is continuous with the endoplasmic reticulum, so the space between the inner and outer nuclear membranes is directly connected with the lumen of the endoplasmic reticulum. In addition, the outer nuclear membrane is functionally similar to the membranes of the endoplasmic reticulum and has ribosomes bound to its cytoplasmic surface. In contrast, the inner nuclear membrane carries unique proteins that are specific to the nucleus.

The critical function of the nuclear membranes is to act as a barrier that separates the contents of the nucleus from the cytoplasm. Like other cell membranes, the nuclear membranes are phospholipid bilayers, which are permeable only to small nonpolar molecules. Other molecules are unable to diffuse through the phospholipid bilayer. The inner and outer nuclear membranes are joined at nuclear pore complexes, the sole channels through which small polar molecules and macromolecules are able to travel through the nuclear envelope

Underlying the inner nuclear membrane is the nuclear lamina, a fibrous meshwork that provides structural support to the nucleus. The nuclear lamina is composed of one or more related proteins called lamins. In addition to providing structural support to the nucleus, the nuclear lamina is thought to serve as a site of chromatin attachment.

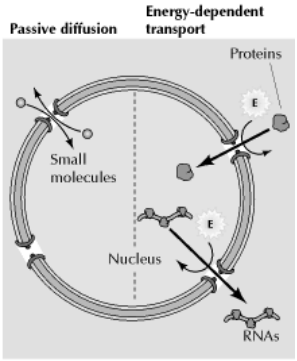
The Nuclear Pore Complex

The nuclear pore complexes are the only channels through which small polar molecules, ions, and macromolecules (proteins and RNAs) are able to travel between the nucleus and the cytoplasm.The nuclear pore complex is an extremely large structure with a diameter of about 120 nm and an estimated molecular mass of approximately 125 million daltons—about 30 times the size of a ribosome. In vertebrates, the nuclear pore complex is composed of 50 to 100 different proteins.



By controlling the traffic of molecules between the nucleus and cytoplasm, the nuclear pore complex plays a fundamental role in the physiology of all eukaryotic cells. RNAs that are synthesized in the nucleus must be efficiently exported to the cytoplasm, where they function in protein synthesis. Conversely, proteins required for nuclear functions (e.g., transcription factors) must be transported into the nucleus from their sites of synthesis in the cytoplasm. In addition, many proteins shuttle continuously between the nucleus and the cytoplasm. The regulated traffic of proteins and RNAs through the nuclear pore complex thus determines the composition of the nucleus and plays a key role in gene expression.

Depending on their size, molecules can travel through the nuclear pore complex by one of two different mechanisms (Figure). Small molecules and some proteins with molecular mass less than approximately 50 kd pass freely across the nuclear envelope in either direction: cytoplasm to nucleus or nucleus to cytoplasm. These molecules diffuse passively through open aqueous channels, estimated to have diameters of approximately 9 nm, in the nuclear pore complex. Most proteins and RNAs, however, are unable to pass through these open channels. Instead, these macromolecules pass through the nuclear pore complex by an active process in which appropriate proteins and RNAs are recognized and selectively transported in only one direction (nucleus to cytoplasm or cytoplasm to nucleus). The traffic of these molecules occurs through regulated channels in the nuclear pore complex that, in response to appropriate signals, can open to a diameter of more than 25 nm a size sufficient to accommodate large ribonucleoprotein complexes, such as ribosomal subunits. It is through these regulated channels that nuclear proteins are selectively imported from the cytoplasm to the nucleus while RNAs are exported from the nucleus to the cytoplasm.



Visualization of nuclear pore complexes by electron microscopy reveals a structure with eightfold symmetry organized around a large central channel, which is the route through which proteins and RNAs cross the nuclear envelope. Detailed structural studies, including computer-based image analysis, have led to the development of three-dimensional models of the nuclear pore complex. These studies indicate that the nuclear pore complex consists of an assembly of eight spokes arranged around a central channel. The spokes are connected to rings at the nuclear and cytoplasmic surfaces, and the spoke-ring assembly is anchored within the nuclear envelope at sites of fusion between the inner and outer nuclear membranes. Protein filaments extend from both the cytoplasmic and nuclear rings, forming a distinct basketlike structure on the nuclear side. The central channel is approximately 40 nm in diameter, which is wide enough to accommodate the largest particles able to cross the nuclear envelope. It contains a structure called the central transporter, through which the active transport of macromolecules is thought to occur.

Internal Organization of the Nucleus

The nucleus is more than a container in which chromatin, RNAs, and nuclear proteins move freely in aqueous solution. Instead, the nucleus appears to have an internal structure that organizes the genetic material and localizes some nuclear functions to discrete sites. The most obvious aspect of the internal organization of the nucleus is the nucleolus, which, as discussed in the following section, is the site at which the rRNA genes are transcribed and ribosomal subunits are assembled. Additional elements of internal nuclear structure are suggested by the organization of chromosomes and by the potential localization of functions such as DNA replication and pre-mRNA processing to distinct nuclear domains.

The Nucleolus

The most prominent substructure within the nucleus is the nucleolus, which is the site of rRNA transcription and processing, and of ribosome assembly. As discussed in the preceding chapter, cells require large numbers of ribosomes to meet their needs for protein synthesis. Actively growing mammalian cells, for example, contain 5 million to 10 million ribosomes that must be synthesized each time the cell divides. The nucleolus is a ribosome production factory, designed to fulfill the need for large-scale production of rRNAs and assembly of the ribosomal subunits.

The Nucleus during Mitosis

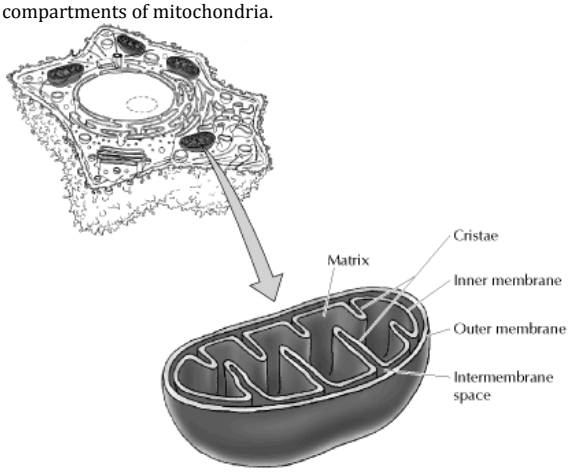
A unique feature of the nucleus is that it disassembles and re-forms each time most cells divide. At the beginning of mitosis, the chromosomes condense, the nucleolus disappears, and the nuclear envelope breaks down, resulting in the release of most of the contents of the nucleus into the cytoplasm. At the end of mitosis, the process is reversed: The chromosomes decondense, and nuclear envelopes re-form around the separated sets of daughter chromosomes. The process is controlled largely by reversible phosphorylation and dephosphorylation of nuclear proteins resulting from the action of the Cdc2 protein kinase, which is a critical regulator of mitosis in all eukaryotic cells.

B3. Mitochondria

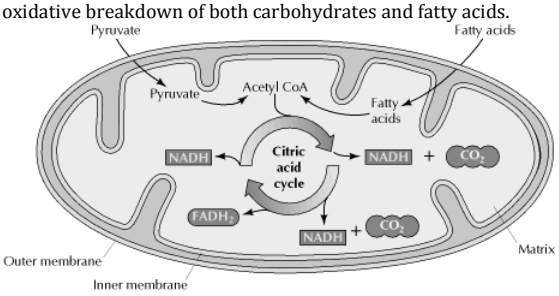
Mitochondria play a critical role in the generation of metabolic energy in eukaryotic cells. They are responsible for most of the useful energy derived from the breakdown of carbohydrates and fatty acids, which is converted to ATP by the process of oxidative phosphorylation. Most mitochondrial proteins are translated on free cytosolic ribosomes and imported into the organelle by specific targeting signals. In addition, mitochondria are unique among the cytoplasmic organelles in that they contain their own DNA, which encodes tRNAs, rRNAs, and some mitochondrial proteins. The assembly of mitochondria thus involves proteins encoded by their own genomes and translated within the organelle, as well as proteins encoded by the nuclear genome and imported from the cytosol.

Organization and Function of Mitochondria

Mitochondria are surrounded by a double-membrane system, consisting of inner and outer mitochondrial membranes separated by an intermembrane space (Figure). The inner membrane forms numerous folds (cristae), which extend into the interior (or matrix) of the organelle. Each of these components plays distinct functional roles, with the matrix and inner membrane representing the major working compartments of mitochondria.

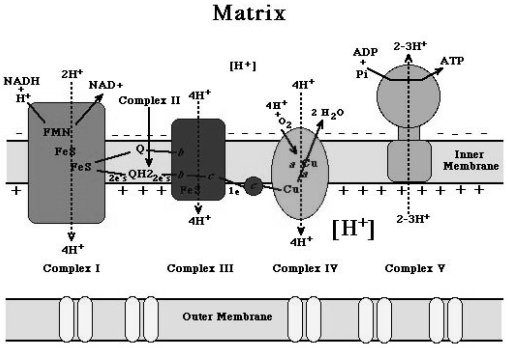


The matrix contains the mitochondrial genetic system as well as the enzymes responsible for the central reactions of oxidative metabolism. The oxidative breakdown of glucose and fatty acids is the principal source of metabolic energy in animal cells. The initial stages of glucose metabolism (glycolysis) occur in the cytosol, where glucose is converted to pyruvate. Pyruvate is then transported into mitochondria, where its complete oxidation to CO_2 yields the bulk of usable energy (ATP) obtained from glucose metabolism. This involves the initial oxidation of pyruvate to acetyl CoA, which is then broken down to CO_2 via the citric acid cycle. The oxidation of fatty acids also yields acetyl CoA, which is similarly metabolized by the citric acid cycle in mitochondria. The enzymes of the citric acid cycle (located in the matrix of mitochondria) thus are central players in the oxidative breakdown of both carbohydrates and fatty acids.



The oxidation of acetyl CoA to CO_2 is coupled to the reduction of NAD^+ and FAD to NADH and $FADH_2$, respectively. Most of the energy derived from oxidative metabolism is then produced by the process of oxidative phosphorylation, which takes place in the inner

mitochondrial membrane. The high-energy electrons from NADH and $FADH_2$ are transferred through a series of carriers in the membrane to molecular oxygen. The energy derived from these electron transfer reactions is converted to potential energy stored in a proton gradient across the membrane, which is then used to drive ATP synthesis. The inner mitochondrial membrane thus represents the principal site of ATP generation, and this critical role is reflected in its structure. First, its surface area is substantially increased by its folding into cristae. In addition, the inner mitochondrial membrane contains an unusually high percentage (greater than 70%) of proteins, which are involved in oxidative phosphorylation as well as in the transport of metabolites (e.g., pyruvate and fatty acids) between the cytosol and mitochondria. Otherwise, the inner membrane is impermeable to most ions and small molecules a property critical to maintaining the proton gradient that drives oxidative phosphorylation.



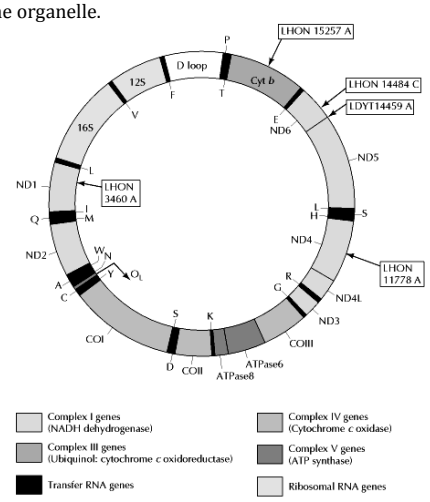
In contrast to the inner membrane, the outer mitochondrial membrane is freely permeable to small molecules. This is because it contains proteins called porins, which form channels that allow the free diffusion of molecules smaller than about 6000 daltons. The composition of the intermembrane space is therefore similar to the cytosol with respect to ions and small molecules. Consequently, the inner mitochondrial membrane is the functional barrier to the passage of small molecules between the cytosol and the matrix and maintains the proton gradient that drives oxidative phosphorylation.

The Genetic System of Mitochondria

Mitochondria contain their own genetic system, which is separate and distinct from the nuclear genome of the cell. Mitochondria are thought to have evolved from bacteria that developed a symbiotic relationship in which they lived within larger cells (endosymbiosis). This hypothesis has recently been substantiated by the results of DNA sequence analysis, which revealed striking similarities between the genomes of mitochondria and of the bacterium *Rickettsia prowazekii*. *Rickettsia* are intracellular parasites which, like mitochondria, are only able to reproduce within eukaryotic cells. Consistent with their similar symbiotic lifestyles, the genomic DNA sequences of *Rickettsia* and mitochondria suggest that they share a common ancestor, from which the genetic system of present-day mitochondria evolved.

Mitochondrial genomes are usually circular DNA molecules, like those of bacteria, which are present in multiple copies per organelle. They vary considerably in size between different species. The genomes of human and most other animal mitochondria are only about 16 kb, but substantially larger mitochondrial genomes are found in yeasts (approximately 80 kb) and plants (more than 200 kb). However, these larger mitochondrial genomes are composed predominantly of noncoding sequences and do not appear to contain significantly more genetic information. For example, the largest sequenced mitochondrial genome is that of the plant *Arabidopsis thaliana*. Although Arabidopsis mitochondrial DNA is approximately 367 kb, it encodes only 32 proteins: just more than twice the number encoded by human mitochondrial DNA. Mitochondrial genomes encode all of the ribosomal RNAs and most of the transfer RNAs needed for translation of these protein-coding sequences within mitochondria. Other mitochondrial proteins are encoded by nuclear genes, which are thought to have been transferred to the nucleus from the ancestral mitochondrial genome.

The human mitochondrial genome encodes 13 proteins involved in electron transport and oxidative phosphorylation (Figure). In addition, human mitochondrial DNA encodes 16S and 12S rRNAs and 22 tRNAs, which are required for translation of the proteins encoded by the organelle genome. The two rRNAs are the only RNA components of animal and yeast mitochondrial ribosomes, in contrast to the three rRNAs of bacterial ribosomes (23S, 16S, and 5S). Plant mitochondrial DNAs, however, also encode a third rRNA of 5S. The mitochondria of plants and protozoans also differ in importing and utilizing tRNAs encoded by the nuclear as well as the mitochondrial genome, whereas in animal mitochondria, all the tRNAs are encoded by the organelle.



The small number of tRNAs encoded by the mitochondrial genome highlights an important feature of the mitochondrial genetic system the use of a slightly different genetic code, which is distinct from the "universal" genetic code used by both prokaryotic and eukaryotic cells . There are 64 possible triplet codons, of which 61 encode the 20 different amino acids incorporated into proteins. Many tRNAs in both prokaryotic and eukaryotic cells are able to recognize more than a single codon in mRNA because of "wobble," which allows some mispairing between the tRNA anticodon and the third position of certain complementary codons.

However, at least 30 different tRNAs are required to translate the universal code according to the wobble rules. Yet human mitochondrial DNA encodes only 22 tRNA species, and these are the only tRNAs used for translation of mitochondrial mRNAs. This is accomplished by an extreme form of wobble in which U in the anticodon of the tRNA can pair with any of the four bases in the third codon position of mRNA, allowing four codons to be recognized by a single tRNA. In addition, some codons specify different amino acids in mitochondria than in the universal code.

Like the DNA of nuclear genomes, mitochondrial DNA can be altered by mutations, which are frequently deleterious to the organelle. Since almost all the mitochondria of fertilized eggs are contributed by the oocyte rather than by the sperm, germ-line mutations in mitochondrial DNA are transmitted

to the next generation by the mother. Such mutations have been associated with a number of diseases. For example, **Leber's hereditary optic neuropathy**, a disease that leads to blindness, can be caused by mutations in mitochondrial genes that encode components of the electron transport chain. In addition, the progressive accumulation of mutations in mitochondrial DNA during the lifetime of individuals has been suggested to contribute to the process of aging.

Mutation in one of the mitochondrial transfer RNA genes, characterized by a decrease in synthesis of mitochondrial protein required for electron transport chain and ATP synthesis causes a disorder of muscles termed as a. **Myoclonic epilepsy and ragged red fiber disease (MERRF).**

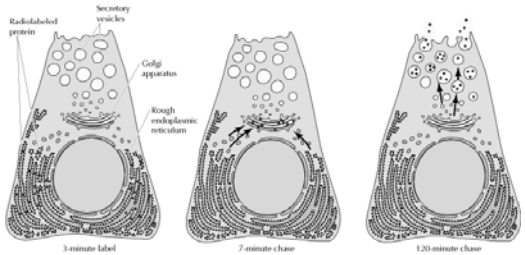
B4. The Endoplasmic Reticulum, Golgi Apparatus, and Lysosomes

In addition to the presence of a nucleus, eukaryotic cells are distinguished from prokaryotic cells by the presence of membrane-enclosed organelles within their cytoplasm. These organelles provide discrete compartments in which specific cellular activities take place, and the resulting subdivision of the cytoplasm allows eukaryotic cells to function efficiently in spite of their large size (about a thousand times the volume of bacteria).

Because of the complex internal organization of eukaryotic cells, the sorting and targeting of proteins to their appropriate destinations are considerable tasks. The first step of protein sorting takes place while translation is still in progress. Many proteins destined for the endoplasmic reticulum, the Golgi apparatus, lysosomes, the plasma membrane, and secretion from the cell are synthesized on ribosomes that are bound to the membrane of the endoplasmic reticulum. As translation proceeds, the polypeptide chains are transported into the endoplasmic reticulum, where protein folding and processing take place. From the endoplasmic reticulum, proteins are transported in vesicles to the Golgi apparatus, where they are further processed and sorted for transport to lysosomes, the plasma membrane, or secretion from the cell. The endoplasmic reticulum, Golgi apparatus, and lysosomes are thus distinguished from other cytoplasmic organelles by their common involvement in protein processing and connection by vesicular transport

The Endoplasmic Reticulum

The endoplasmic reticulum (ER) is a network of membrane-enclosed tubules and sacs (cisternae) that extends from the nuclear membrane throughout the cytoplasm (Figure 9.1). The entire endoplasmic reticulum is enclosed by a continuous membrane and is the largest organelle of most eukaryotic cells. Its membrane may account for about half of all cell membranes, and the space enclosed by the ER (the lumen, or cisternal space) may represent about 10% of the total cell volume. As discussed below, there are two distinct types of ER that perform different functions within the cell. The rough ER, which is covered by ribosomes on its outer surface, functions in protein processing. The smooth ER is not associated with ribosomes and is involved in lipid, rather than protein, metabolism.



The Endoplasmic Reticulum and Protein Secretion

The role of the endoplasmic reticulum in protein processing and sorting was first demonstrated by George Palade and his colleagues in the 1960s (Figure). These investigators studied the fate of newly synthesized proteins in specialized cells of the pancreas (pancreatic acinar cells) that secrete digestive enzymes into the small intestine. Because most proteins synthesized by these cells are secreted, Palade and coworkers were able to study the pathway taken by secreted proteins simply by labeling newly synthesized proteins with radioactive amino acids. The location of the radiolabeled proteins within the cell was then determined by autoradiography, revealing the cellular sites involved in the events leading to protein secretion. After a brief exposure of pancreatic acinar cells to radioactive amino acids, newly synthesized proteins were detected in the rough ER, which was therefore identified as the site of synthesis of proteins destined for secretion. If the cells were then incubated for a short time in media containing nonradioactive amino acids (a process known as a chase), the radiolabeled proteins were detected in the Golgi apparatus. Following longer chase periods, the radiolabeled proteins traveled from the Golgi apparatus to the cell surface in secretory vesicles, which then fused with the plasma membrane to release their contents outside of the cell.

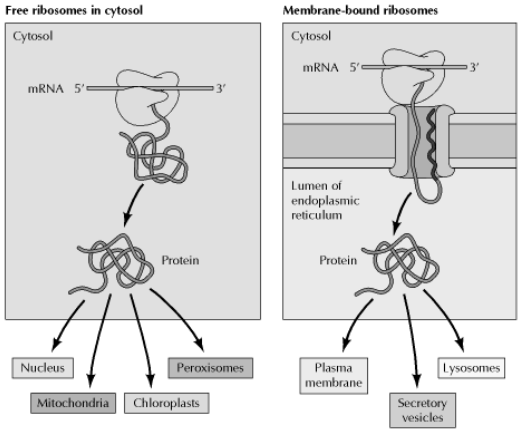
These experiments defined a pathway taken by secreted proteins, the secretory pathway: rough ER → Golgi → secretory vesicles → cell exterior. Further studies extended these results and demonstrated that this pathway is not restricted to proteins destined for secretion from the cell. Plasma membrane and lysosomal proteins also travel from the rough ER to the Golgi and then to their final destinations. Still other proteins travel through the initial steps of the secretory pathway but are then retained and function within either the ER or the Golgi apparatus.

The entrance of proteins into the ER thus represents a major branch point for the traffic of proteins within eukaryotic cells. Proteins destined for secretion or incorporation into the ER, Golgi apparatus, lysosomes, or plasma membrane are initially targeted to the ER. In mammalian cells, most proteins are transferred into the ER while they are being translated on membrane-bound ribosomes (Figure). In contrast, proteins destined to remain in the cytosol or to be incorporated into the nucleus, mitochondria, chloroplasts, or peroxisomes are synthesized on free ribosomes and released into the cytosol when their translation is complete.

Targeting Proteins to the Endoplasmic Reticulum

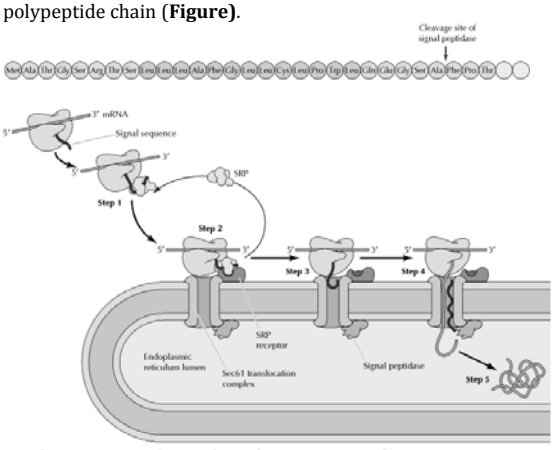
Proteins can be translocated into the ER either during their synthesis on membrane-bound ribosomes (cotranslational translocation) or after their translation has been completed on free ribosomes in the cytosol (posttranslational translocation). In mammalian cells, most proteins enter the ER co-translationally, whereas both cotranslational and posttranslational pathways are used in yeast. The first step in the cotranslational pathway is the association of ribosomes with the ER. Ribosomes are targeted for binding to the ER membrane by the amino acid sequence of the polypeptide chain being synthesized, rather than by intrinsic properties of the ribosome itself. Free and membrane-bound ribosomes are functionally indistinguishable, and all protein

synthesis initiates on ribosomes that are free in the cytosol. Ribosomes engaged in the synthesis of proteins that are destined for secretion are then targeted to the endoplasmic reticulum by a signal sequence at the amino terminus of the growing polypeptide chain. These signal sequences are short stretches of hydrophobic amino acids that are cleaved from the polypeptide chain during its transfer into the ER lumen.



David Sabatini and Günter Blobel first proposed in 1971 that the signal for ribosome attachment to the ER was an amino acid sequence near the amino terminus of the growing polypeptide chain. This hypothesis was supported by the results of in vitro translation of mRNAs encoding secreted proteins, such as immunoglobulins. If an mRNA encoding a secreted protein was translated on free ribosomes in vitro, it was found that the protein produced was slightly larger than the normal secreted protein. If microsomes were added to the system, however, the in vitro-translated protein was incorporated into the microsomes and cleaved to the correct size. These experiments led to a more detailed formulation of the signal hypothesis, which proposed that an amino-terminal leader sequence targets the polypeptide chain to the microsomes and is then cleaved by a microsomal protease. Many subsequent findings have substantiated this model, including recombinant DNA experiments demonstrating that addition of a signal sequence to a normally nonsecreted protein is sufficient to direct the incorporation of the recombinant protein into the rough ER.

The mechanism by which secretory proteins are targeted to the ER during their translation (the cotranslational pathway) is now well understood. The signal sequences span about 20 amino acids, including a stretch of hydrophobic residues, usually at the amino terminus of the polypeptide chain (**Figure)**.

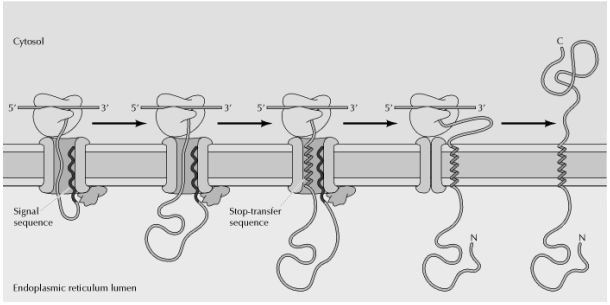


As they emerge from the ribosome, signal sequences are recognized and bound by a **signal recognition particle** (SRP) consisting of six polypeptides and a small cytoplasmic RNA (7SL RNA). SRP binds the ribosome as well as the signal sequence, inhibiting further translation and targeting the entire complex (the SRP, ribosome, and growing polypeptide chain) to the rough ER by binding to the SRP receptor on the ER membrane (Figure). Binding to the receptor releases the SRP from both the ribosome and the signal sequence of the growing polypeptide chain. The ribosome then binds to a protein translocation complex in the ER membrane, and the signal sequence is inserted into a membrane channel. In both yeast and mammalian cells, the translocation channels through the ER membrane are complexes of three transmembrane proteins, called the Sec61 proteins. Thus, the process of protein synthesis directly drives the transfer of growing polypeptide chains through the Sec61 channel and into the ER. As translocation proceeds, the signal sequence is cleaved by signal peptidase and the polypeptide is released into the lumen of the ER.

Insertion of Proteins into the ER Membrane

Proteins destined for secretion or residence within the lumen of the ER, Golgi apparatus, or lysosomes are translocated across the ER membrane and released into the lumen of the ER as already described. However, proteins destined for incorporation into the plasma membrane or the membranes of the ER, Golgi, or lysosomes are initially inserted into the ER membrane instead of being released into the lumen. From the ER membrane, they proceed to their final destination along the same pathway as that of secretory proteins: ER → Golgi → plasma membrane or lysosomes. These proteins are transported along this pathway as membrane components, however, rather than as soluble proteins.

Integral membrane proteins are embedded in the membrane by hydrophobic regions that span the phospholipid bilayer. The membrane-spanning portions of these proteins are usually a-helical regions consisting of 20 to 25 hydrophobic amino acids. The formation of an α- helix maximizes hydrogen bonding between the peptide bonds, and the hydrophobic amino acid side chains interact with the fatty acid tails of the phospholipids. However, different integral membrane proteins differ in how they are inserted. For example, whereas some integral membrane proteins span the membrane only once, others have multiple membrane-spanning regions. In addition, some proteins are oriented in the membrane with their amino terminus on the cytosolic side; others have their carboxy terminus exposed to the cytosol. These orientations of proteins inserted into the ER, Golgi, lysosomal, and plasma membranes are established as the growing polypeptide chains are translocated into the ER. The lumen of the ER is topologically equivalent to the exterior of the cell, so the domains of plasma membrane proteins that are exposed on the cell surface correspond to the regions of polypeptide chains that are translocated into the ER**.**

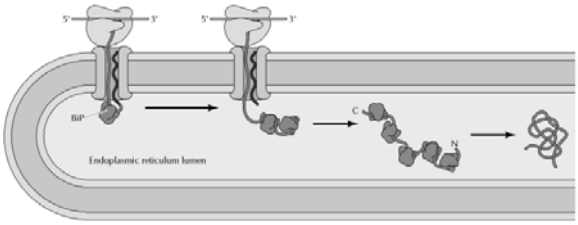


The most straightforward mode of insertion into the ER membrane results in the synthesis of transmembrane proteins oriented with their carboxy termini exposed to the cytosol (Figure). These proteins have a normal amino-

terminal signal sequence, which is cleaved by signal peptidase during translocation of the polypeptide chain across the ER membrane through the Sec61 channel. They are then anchored in the membrane by a second membrane-spanning a helix in the middle of the protein. This transmembrane sequence, called a stop-transfer sequence, signals closure of the Sec61 channel. Further translocation of the polypeptide chain across the ER membrane is thus blocked, so the carboxy-terminal portion of the growing polypeptide chain is synthesized in the cytosol. The transmembrane domain then exits the translocation channel laterally to enter the lipid bilayer. The insertion of these proteins in the membrane thus involves the sequential action of two distinct elements: a cleavable amino-terminal signal sequence that initiates translocation across the membrane and a transmembrane stop-transfer sequence that anchors the protein in the membrane.

Proteins can also be anchored in the ER membrane by internal signal sequences that are not cleaved by signal peptidase. Proteins that span the membrane multiple times are thought to be inserted as a result of an alternating series of internal signal sequences and transmembrane stop-transfer sequences.

Protein Folding and Processing in the ER



For proteins that enter the secretory pathway, many of these events occur either during translocation across the ER membrane or within the ER lumen. One such processing event is the proteolytic cleavage of the signal peptide as the polypeptide chain is translocated across the ER membrane. The ER is also the site of protein folding, assembly of multisubunit proteins, disulfide bond formation, the initial stages of glycosylation, and the addition of glycolipid anchors to some plasma membrane proteins. Indeed, the primary role of lumenal ER proteins is to catalyze the folding and assembly of newly translocated polypeptides.
As already discussed, proteins are translocated across the ER membrane as unfolded polypeptide chains while their translation is still in progress. These polypeptides, therefore, fold into their three-dimensional conformations within the ER, assisted by the molecular chaperones that facilitate the folding of polypeptide chains. For example, one of the major proteins within the ER lumen is a member of the Hsp70 family of chaperones called BiP. BiP is thought to bind to the unfolded polypeptide chain as it crosses the membrane and then mediates protein folding and the assembly of multisubunit proteins within the ER (Figure). Correctly assembled proteins are released from BiP and are available for transport to the Golgi apparatus. Abnormally folded or improperly assembled proteins, however, remain bound to BiP and are consequently retained within the ER or degraded, rather than being transported farther along the secretory pathway.

The formation of disulfide bonds between the side chains of cysteine residues is an important aspect of protein folding and assembly within the ER. Disulfide bond formation is facilitated by the enzyme protein disulfide isomerase which is located in the ER lumen.

Proteins are also glycosylated on specific asparagine residues (**N-linked glycosylation**) within the ER while their translation is still in process.Oligosaccharide units consisting of 14 sugar residues are added to acceptor asparagine residues of growing polypeptide chains as they are translocated into the ER. The oligosaccharide is synthesized on a lipid (dolichol) carrier anchored in the ER membrane. It is then transferred as a unit to acceptor asparagine residues in the consensus sequence **Asn-X-Ser/Thr** by a membrane-bound enzyme called oligosaccharyl transferase.

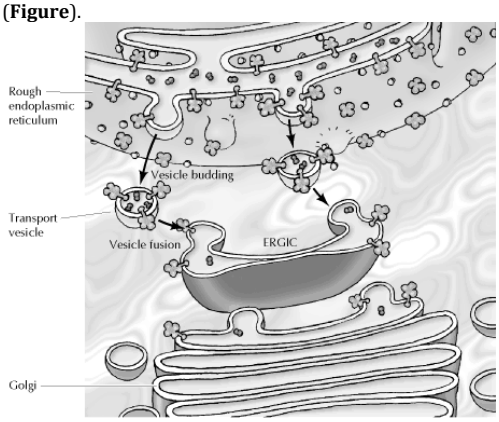
The Smooth ER and Lipid Synthesis

In addition to its activities in the processing of secreted and membrane proteins, the ER is the major site at which membrane lipids are synthesized in eukaryotic cells. Because they are extremely hydrophobic, lipids are synthesized in association with already existing cellular membranes rather than in the aqueous environment of the cytosol. Although some lipids are synthesized in association with other membranes, most are synthesized in the ER. They are then transported from the ER to their ultimate destinations either in vesicles or by carrier proteins.

Smooth ER is abundant in cell types that are particularly active in lipid metabolism. For example, steroid hormones are synthesized (from cholesterol) in the ER, so large amounts of smooth ER are found in steroid-producing cells, such as those in the testis and ovary. In addition, smooth ER is abundant in the liver, where it contains enzymes that metabolize various lipid-soluble compounds. These detoxifying enzymes inactivate a number of potentially harmful drugs (e.g., phenobarbital) by converting them to water-soluble compounds that can be eliminated from the body in the urine. The smooth ER is thus involved in multiple aspects of the metabolism of lipids and lipid-soluble compounds.

Export of Proteins and Lipids from the ER

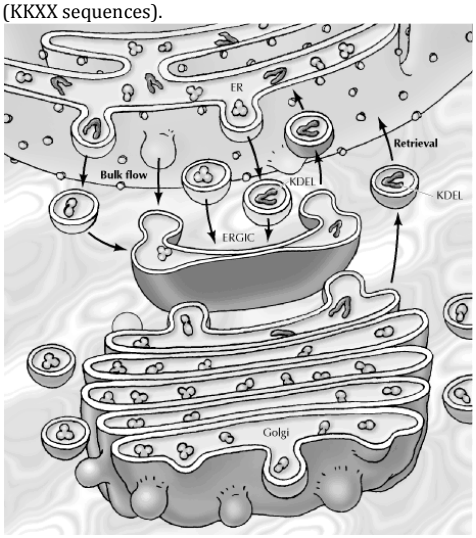
Both proteins and lipids travel along the secretory pathway in transport vesicles, which bud from the membrane of one organelle and then fuse with the membrane of another. Thus, molecules are exported from the ER in vesicles that bud from the ER and carry their cargo first to the ER-Golgi intermediate compartment and then to the Golgi apparatus (**Figure**).



Subsequent steps in the secretory pathway involve vesicular transport between different compartments of the Golgi and from the Golgi to lysosomes or the plasma membrane. In each case, proteins within the lumen of one organelle are packaged into the budding transport vesicle and then released into the lumen of the recipient organelle following vesicle fusion. Membrane proteins and lipids are transported similarly, and it is noteworthy that their topological orientation is maintained as they travel from one membrane-enclosed organelle to another. For example, the

domains of a protein exposed on the cytosolic side of the ER membrane will also be exposed on the cytosolic side of the Golgi and plasma membranes, whereas protein domains exposed on the lumenal side of the ER membrane will be exposed on the lumenal side of the Golgi and on the exterior of the cell.

While most proteins travel from the ER to the Golgi, some proteins must be retained within the ER rather than proceeding along the secretory pathway. In particular, proteins that function within the ER (including BiP, signal peptidase, protein disulfide isomerase, and other enzymes discussed earlier) must be retained within that organelle. The distinction between proteins exported from and those retained in the ER appears to be governed by two distinct types of targeting sequences that specifically mark proteins as either (1) destined for transport to the Golgi or (2) destined for retention in the ER. Many proteins are retained in the ER lumen as a result of the presence of the targeting sequence Lys-Asp-Glu-Leu (**KDEL**, in the single-letter code) at their carboxy terminus.The retention of some transmembrane proteins in the ER is similarly dictated by short C-terminal sequences that contain two lysine residues (KKXX sequences).



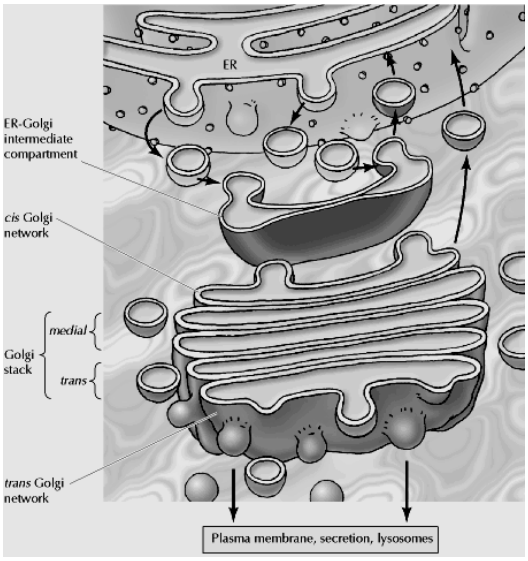
The Golgi Apparatus

The Golgi apparatus, or Golgi complex, functions as a factory in which proteins received from the ER are further processed and sorted for transport to their eventual destinations: lysosomes, the plasma membrane, or secretion. In addition, as noted earlier, glycolipids and sphingomyelin are synthesized within the Golgi. In plant cells, the Golgi apparatus further serves as the site at which the complex polysaccharides of the cell wall are synthesized. The Golgi apparatus is thus involved in processing the broad range of cellular constituents that travel along the secretory pathway.

Organization of the Golgi

Morphologically the Golgi is composed of flattened membrane-enclosed sacs (cisternae) and associated vesicles. A striking feature of the Golgi apparatus is its distinct polarity in both structure and function. Proteins from the ER enter at its cis face (entry face), which is convex and usually oriented toward the nucleus. They are then transported through the Golgi and exit from its concave trans face (exit face). As they pass through the Golgi, proteins are modified and sorted for transport to their eventual destinations within the cell.

Distinct processing and sorting events appear to take place in an ordered sequence within different regions of the Golgi complex, so the Golgi is usually considered to consist of multiple discrete compartments. Although the number of such compartments has not been established, the Golgi is most commonly viewed as consisting of four functionally distinct regions: the cis-Golgi network, the Golgi stack (which is divided into the medial and trans subcompartments), and the transGolgi network (Figure).



Proteins from the ER are transported to the ER-Golgi intermediate compartment and then enter the Golgi apparatus at the cis Golgi network. They then progress to the medial and trans compartments of the Golgi stack, within which most metabolic activities of the Golgi apparatus take place. The modified proteins, lipids, and polysaccharides then move to the trans Golgi network, which acts as a sorting and distribution center, directing molecular traffic to lysosomes, the plasma membrane, or the cell exterior.

Protein Glycosylation within the Golgi

Protein processing within the Golgi involves the modification and synthesis of the carbohydrate portions of glycoproteins. One of the major aspects of this processing is the modification of the N-linked oligosaccharides that were added to proteins in the ER. As discussed earlier in this chapter, proteins are modified within the ER by the addition of an oligosaccharide consisting of 14 sugar residues. Three glucose residues and one mannose are then removed while the polypeptides are still in the ER. Following transport to the Golgi apparatus, the N-linked oligosaccharides of these glycoproteins are subject to extensive further modifications.

Proteins can also be modified by the addition of carbohydrates to the side chains of acceptor serine and threonine residues within specific sequences of amino acids **(O-linked glycosylation).** These modifications take place in the Golgi apparatus by the sequential addition of single sugar residues. The serine or threonine is usually linked directly to N-acetylgalactosamine, to which other sugars can then be added.Proteins destined for incorporation into lysosomes are linked by mannose 6-phosphate

Lipid and Polysaccharide Metabolism in the Golgi

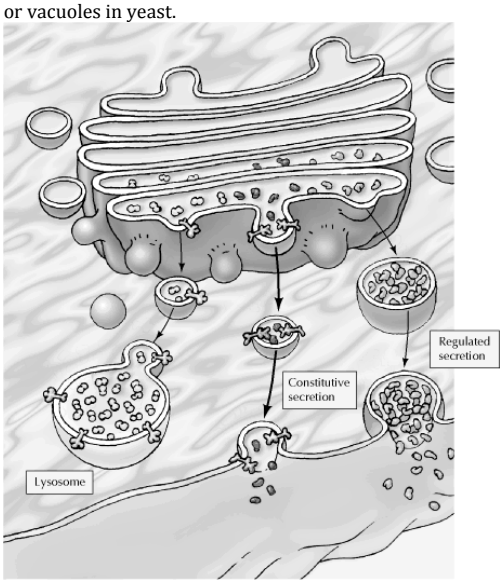
In addition to its activities in processing and sorting glycoproteins, the Golgi apparatus functions in lipid metabolism- in particular, in the synthesis of glycolipids and

sphingomyelin. As discussed earlier, the glycerol phospholipids, cholesterol, and ceramide are synthesized in the ER. Sphingomyelin and glycolipids are then synthesized from ceramide in the Golgi apparatus. Sphingomyelin (the only nonglycerol phospholipid in cell membranes) is synthesized by the transfer of a phosphorylcholine group from phosphatidylcholine to ceramide. Alternatively, the addition of carbohydrates to ceramide can yield a variety of different glycolipids.

In plant cells, the Golgi apparatus has the additional task of serving as the site where complex polysaccharides of the cell wall are synthesized. The plant cell wall is composed of three major types of polysaccharides. Cellulose, the predominant constituent, is a simple linear polymer of glucose residues. It is synthesized at the cell surface by enzymes in the plasma membrane. The other cell wall polysaccharides (hemicelluloses and pectins), however, are complex, branched chain molecules that are synthesized in the Golgi apparatus and then transported in vesicles to the cell surface. The synthesis of these cell wall polysaccharides is a major cellular function, and as much as 80% of the metabolic activity of the Golgi apparatus in plant cells may be devoted to polysaccharide synthesis.

Protein Sorting and Export from the Golgi Apparatus

Proteins, as well as lipids and polysaccharides, are transported from the Golgi apparatus to their final destinations through the secretory pathway. This involves the sorting of proteins into different kinds of transport vesicles, which bud from the trans Golgi network and deliver their contents to the appropriate cellular locations (Figure). Some proteins are carried from the Golgi to the plasma membrane by a constitutive secretory pathway, which accounts for the incorporation of new proteins and lipids into the plasma membrane, as well as for the continuous secretion of proteins from the cell. Other proteins are transported to the cell surface by a distinct pathway of regulated secretion or are specifically targeted to other intracellular destinations, such as lysosomes in animal cells or vacuoles in yeast.



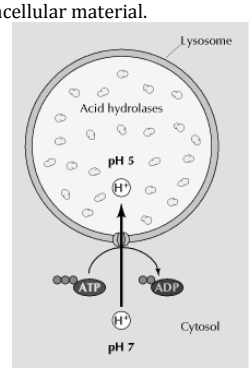
Proteins that function within the Golgi apparatus must be retained within that organelle, rather than being transported along the secretory pathway. In contrast to the ER, all of the proteins retained within the Golgi complex are associated with the Golgi membrane rather than being soluble proteins within the lumen. The signals responsible for retention of some proteins within the Golgi have been localized to their transmembrane domains, which retain proteins within the Golgi apparatus by preventing them from being packaged in the transport vesicles that leave the trans Golgi network. In addition, like the KKXX sequences of resident ER membrane proteins, signals in the cytoplasmic tails of some Golgi proteins mediate the retrieval of these proteins from subsequent compartments along the secretory pathway.

The constitutive secretory pathway, which operates in all cells, leads to continual unregulated protein secretion. However, some cells also possess a distinct regulated secretory pathway in which specific proteins are secreted in response to environmental signals. Examples of regulated secretion include the release of hormones from endocrine cells, the release of neurotransmitters from neurons, and the release of digestive enzymes from the pancreatic acinar cells. Proteins are sorted into the regulated secretory pathway in the trans Golgi network, where they are packaged into specialized secretory vesicles. These secretory vesicles, which are larger than other transport vesicles, store their contents until specific signals direct their fusion with the plasma membrane. For example, the digestive enzymes produced by pancreatic acinar cells are stored in secretory vesicles until the presence of food in the stomach and small intestine triggers their secretion. The sorting of proteins into the regulated secretory pathway appears to involve the recognition of signal patches shared by multiple proteins that enter this pathway. These proteins selectively aggregate in the trans Golgi network and are then released by budding as secretory vesicles.

The best-characterized pathway of protein sorting in the Golgi is the selective transport of proteins to lysosomes. As already discussed, lumenal lysosomal proteins are marked by mannose-6-phosphates that are formed by modification of their N-linked oligosaccharides shortly after entry into the Golgi apparatus. A specific receptor in the membrane of the trans Golgi network then recognizes these mannose-6-phosphate residues. The resulting complexes of receptor plus lysosomal enzyme are packaged into transport vesicles destined for lysosomes. Lysosomal membrane proteins are targeted by sequences in their cytoplasmic tails, rather than by mannose-6-phosphates.

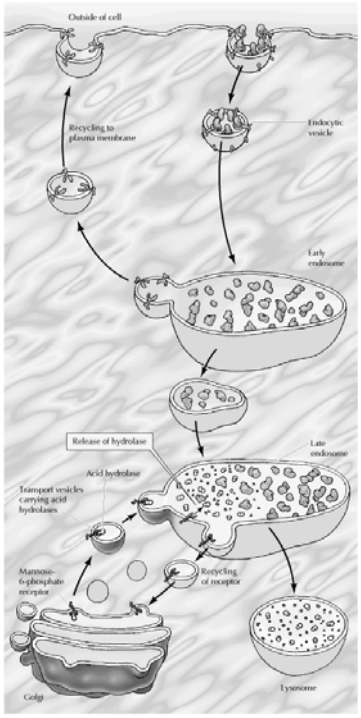
Lysosomes

Lysosomes are membrane-enclosed organelles that contain an array of enzymes capable of breaking down all types of biological polymers—proteins, nucleic acids, carbohydrates, and lipids. Lysosomes function as the digestive system of the cell, serving both to degrade material taken up from outside the cell and to digest obsolete components of the cell itself. In their simplest form, lysosomes are visualized as dense spherical vacuoles, but they can display considerable variation in size and shape as a result of differences in the materials that have been taken up for digestion. Lysosomes thus represent morphologically diverse organelles defined by the common function of degrading intracellular material.



Lysosomes contain about 50 different degradative enzymes that can hydrolyze proteins, DNA, RNA, polysaccharides, and lipids. Mutations in the genes that encode these enzymes are responsible for more than 30 different human genetic diseases, which are called lysosomal storage diseases because undegraded material accumulates within the lysosomes of affected individuals. Most of these diseases result from deficiencies in single lysosomal enzymes. For example, **Gaucher's disease** (the most common of these disorders) results from a mutation in the gene that encodes a lysosomal enzyme required for the breakdown of glycolipids. An intriguing exception is I-cell disease, which is caused by a deficiency in the enzyme that catalyzes the first step in the tagging of lysosomal enzymes with mannose-6-phosphate in the Golgi apparatus. The result is a general failure of lysosomal enzymes to be incorporated into lysosomes.

All of the lysosomal enzymes are acid hydrolases, which are active at the acidic pH (about 5) that is maintained within lysosomes but not at the neutral pH (about 7.2) characteristic of the rest of the cytoplasm (Figure). The requirement of these lysosomal hydrolases for acidic pH provides double protection against uncontrolled digestion of the contents of the cytosol; even if the lysosomal membrane were to break down, the released acid hydrolases would be inactive at the neutral pH of the cytosol. To maintain their acidic internal pH, lysosomes must actively concentrate H+ ions (protons). This is accomplished by a proton pump in the lysosomal membrane, which actively transports protons into the lysosome from the cytosol. This pumping requires expenditure of energy in the form of ATP hydrolysis, since it maintains approximately a hundredfold higher H+ concentration inside the lysosome.



Endocytosis and Lysosome Formation

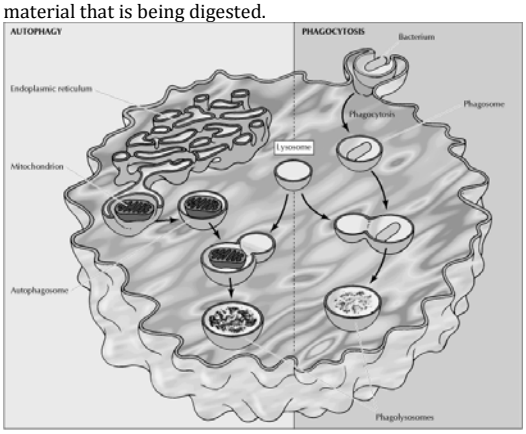
One of the major functions of lysosomes is the digestion of material taken up from outside the cell by endocytosis.. However, the role of lysosomes in the digestion of material taken up by endocytosis relates not only to the function of lysosomes but also to their formation. In particular, lysosomes are formed by the fusion of transport vesicles budded from the trans Golgi network with endosomes, which contain molecules taken up by endocytosis at the plasma membrane.

The formation of lysosomes thus represents an intersection between the secretory pathway, through which lysosomal proteins are processed, and the endocytic pathway, through which extracellular molecules are taken up at the cell surface (Figure). Material from outside the cell is taken up in clathrin-coated endocytic vesicles, which bud from the plasma membrane and then fuse with early endosomes. Membrane components are then recycled to the plasma membrane and the early endosomes gradually mature into late endosomes, which are the precursors to lysosomes. One of the important changes during endosome maturation is the lowering of the internal pH to about 5.5, which plays a key role in the delivery of lysosomal acid hydrolases from the trans Golgi network.

As discussed earlier, acid hydrolases are targeted to lysosomes by mannose-6-phosphate residues, which are recognized by mannose-6-phosphate receptors in the trans Golgi network and packaged into clathrin-coated vesicles. Following removal of the clathrin coat, these transport vesicles fuse with late endosomes, and the acidic internal pH causes the hydrolases to dissociate from the mannose-6-phosphate receptor. The hydrolases are thus released into the lumen of the endosome, while the receptors remain in the membrane and are eventually recycled to the Golgi. Late endosomes then mature into lysosomes as they acquire a full complement of acid hydrolases, which digest the molecules originally taken up by endocytosis.

Phagocytosis and Autophagy

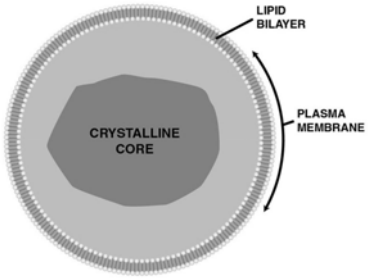
In addition to degrading molecules taken up by endocytosis, lysosomes digest material derived from two other routes: phagocytosis and autophagy (Figure). In phagocytosis, specialized cells, such as macrophages, take up and degrade large particles, including bacteria, cell debris, and aged cells that need to be eliminated from the body. Such large particles are taken up in phagocytic vacuoles (phagosomes), which then fuse with lysosomes, resulting in digestion of their cont ents. The lysosomes formed in this way (phagolysosomes) can be quite large and heterogeneous, since their size and shape is determined by the content of material that is being digested.



Lysosomes are also responsible for autophagy, the gradual turnover of the cell's own components. The first step of autophagy appears to be the enclosure of an organelle (e.g., a mitochondrion) in membrane derived from the ER. The resulting vesicle (an autophagosome) then fuses with a lysosome, and its contents are digested.

B5. Peroxisomes

Peroxisomes are ubiquitous organelles in eukaryotes that participate in the metabolism of fatty acids and other metabolites. Peroxisomes have enzymes that rid the cell of toxic peroxides. They have a single lipid bilayer membrane that separates their contents from the cytosol (the internal fluid of the cell) and contain membrane proteins critical for various functions, such as importing proteins into the organelles and aiding in proliferation. Like lysosomes, peroxisomes are part of the secretory pathway of a cell, but they are much more dynamic and can replicate by enlarging and then dividing. Peroxisomes were identified as cellular organelles by the Belgian cytologist Christian de Duve in 1967 after they had been first described in a Swedish PhD thesis a decade earlier.



Occurrence and evolution: Peroxisomes are found in virtually all eukaryotic cells. Peroxisomes contain enzymes for certain oxidative reactions, like the beta-oxidation of very-long-chain fatty acids. Prokaryotes lack peroxisomes. The enzymatic content of peroxisomes varies across species, but the presence of certain proteins common to many species has been used to suggest an endosymbiotic origin; that is, peroxisomes evolved from bacteria that invaded larger cells as parasites, and very gradually evolved a symbiotic relationship. However, this view has been challenged by recent discoveries. For example, peroxisome-less mutants can restore peroxisomes upon introduction of the wild-type gene, and peroxisomes have been observed to be formed from the endoplasmic reticulum.

An evolutionary analysis of the peroxisomal proteome found homologies between the peroxisomal import machinery and the ERAD pathway in the endoplasmic reticulum, along with a number of metabolic enzymes that were likely recruited from the mitochondria.[6] These results indicate that the peroxisome does not have an endosymbiotic origin; instead, it likely originates from the ER, and its proteins were recruited from pools existing within the primitive eukaryote, as quoted in the science textbook Biozone.

Function: Peroxisomes contain oxidative enzymes, such as catalase, D-amino acid oxidase, and uric acid oxidase. Certain enzymes within the peroxisome, by using molecular oxygen, remove hydrogen atoms from specific organic substrates (labeled as R), in an oxidative reaction, producing hydrogen peroxide (H_2O_2, itself toxic):

$$RH_2+O_2 \rightarrow R+H_2O_2$$

Catalase, another enzyme in the peroxisome, in turn uses this H2O2 to oxidize other substrates, including phenols, formic acid, formaldehyde, and alcohol, by means of the peroxidation reaction: $H_2\ O_2+\ R'H_2 \rightarrow R'\ +\ 2H_2O$, thus eliminating the poisonous hydrogen peroxide in the process.

This reaction is important in liver and kidney cells, where the peroxisomes detoxify various toxic substances that enter the blood. About 25% of the ethanol we drink is oxidized to acetaldehyde in this way. In addition, when excess H2O2 accumulates in the cell, catalase converts it to H_2O through this reaction:

$$2H_2O_2 \rightarrow 2H_2O+O_2$$

A major function of the peroxisome is the breakdown of fatty acid molecules, in a process called beta-oxidation. In this process, the fatty acids are broken down two carbons at a time, converted to Acetyl-CoA, which is then transported back to the cytosol for further use. In animal cells, beta-oxidation can also occur in the mitochondria. In yeast and plant cells, this process is exclusive for the peroxisome.

The first reactions in the formation of plasmalogen in animal cells also occurs in peroxisomes. Plasmalogen is the most abundant phospholipid in myelin. Deficiency of plasmalogens causes profound abnormalities in the myelination of nerve cells, which is one of the reasons that many peroxisomal disorders lead to neurological disease.

Peroxisomes also play a role in the production of bile acids and proteins.

In higher plants, peroxisomes contain also a complex battery of antioxidative enzymes such as superoxide dismutase, the components of the ascorbate-glutathione cycle, and the NADP-dehydrogenases of the pentose-phosphate pathway. It has been demonstrated the generation of superoxide ($O_2\bullet$-) and nitric oxide (•NO) radicals. They are also involved in photorespiration.

In plants special type of peroxisomes termed as glyoxysome carry out glyoxylate cycle in which lipids stored in seed are converted to carbohydrates and thus provides source of nutrition for developing embryo.

Protein import: Proteins are selectively imported into peroxisomes. Since the organelles contain no DNA or ribosomes and thus have no means of producing proteins, all of their proteins must be imported across the membrane. It is believed that necessary proteins enter through the endoplasmic reticulum during biogenesis as well as through membrane proteins.

A specific protein signal (PTS or peroxisomal targeting signal) of three amino acids at the C-terminus of many peroxisomal proteins signals the membrane of the peroxisome to import them into the organelle. Other peroxisomal proteins contain a signal at the N-terminus. There are at least 32 known peroxisomal proteins, called peroxins, which participate in the process of importing proteins by means of ATP hydrolysis. Proteins do not have to unfold to be imported into the peroxisome. The protein receptors, the peroxins PEX5 and PEX7, accompany their cargoes (containing a PTS1 or a PTS2, respectively) all the way into the peroxisome where they release the cargo and then return to the cytosol - a step named recycling. Overall, the import cycle is referred to as the extended shuttle mechanism. Evidence now indicates that ATP hydrolysis is required for the recycling of receptors to the cytosol. Also, ubiquitination appears to be crucial for the export of PEX5 from the peroxisome, to the cytosol. Little is known about the import of PEX7, although it has helper proteins that have been shown to be ubiquitinated.

Deficiencies: Peroxisomal disorders are a class of condtions that lead to disorders of lipid metabolism. One well-known example is Zellweger syndrome. Peroxisomes matrix proteins are synthesized on free ribosomes in the cytosol and that these proteins are imported posttranslationally in pre-existing peroxisomes.

B6. Vacuoles

Vacuoles are found in the cytoplasm of most plant cells and some animal cells. Vacuoles are membrane-bound compartments within some eukaryotic cells that can serve a variety of secretory, excretory, and storage functions.

Vacuoles and their contents are considered to be distinct from the cytoplasm, and are classified as ergastic according to some authors. Vacuoles are especially conspicuous in most plant cells.

Functions:

In general, vacuole functions include also

- Removing unwanted structual debris
- Isolating materials that might be harmful or a threat to the cell
- Containing waste products
- Maintaining internal hydrostatic pressure or turgor within the cell
- Maintaining an acidic internal pH
- Containing small molecules
- Exporting unwanted substances from the cell

Vacuoles also play a major role in autophagy, maintaining a balance between biogenesis (production) and degradation (or turnover), of many substances and cell structures. Vacuoles store food and other materials needed by a cell. They also aid in destruction of invading bacteria or of misfolded proteins that have begun to build up within the cell. The vacuole is a major part in the plant and animal cell.

Protists:

Some protists and macrophages use food vacuoles as a stage in phagocytosis—the intake of large molecules, particles, or even other cells, by the cell for digestion. They are also called "storage sacs."

A contractile vacuole is used to pump excess water out of the cell to reduce osmotic pressure and keep the cell from bursting, which is referred to as cytolysis or osmotic lysis.

Budding yeast:

In budding yeast cells, vacuoles act as storage compartments of amino acids and detoxification compartments. Under conditions of starvation, proteins are degraded in vacuoles; this is called autophagy. First, cytoplasms, mitochondrion, and small organelles are covered with multiplex plasma membranes called autophagosomes. Next, the autophagosomes fuse the vacuoles. Finally, the cytoplasms and the organelles are degraded.

In a vacuole of budding yeast, black particles sometimes appear, called a dancing body. The dancing body moves actively in the vacuole and appears and disappears within 10 minutes to several hours. In previous research, it was suggested but not proven that the main component of the dancing body is polyphosphate acid. But the main component has been determined to be crystallized sodium polyphosphate and its function has been studied. It is thought that its function is to supply and store phosphates in budding yeast cells.

Plants:

Most mature plant cells have one or several vacuoles that typically occupy more than 30% of the cell's volume, and that can occupy as much as 90% of the volume for certain cell types and conditions. A vacuole is surrounded by a membrane called the tonoplast.

This vacuole houses large amounts of a liquid called cell sap, composed of water, enzymes, inorganic ions (like K^+ and Cl^-), salts (such as calcium), and other substances, including toxic byproducts removed from the cytosol to avoid interference with metabolism.

Toxins present in the vacuole may also help to protect some plants from predators.

Transport of protons from cytosol to vacuole aids in keeping cytoplasmic pH stable, while making the vacuolar interior more acidic, allowing degradative enzymes to act.

Although having a large central vacuole is the most common case, the size and number of vacuoles may vary in different tissues and stages of development. Cells of the vascular cambium, for example, have many small vacuoles in winter, and one large one in summer.

Aside from storage, the main role of the central vacuole is to maintain turgor pressure against the cell wall. Proteins found in the tonoplast control the flow of water into and out of the vacuole through active transport, pumping potassium (K^+) ions into and out of the vacuolar interior.

Due to osmosis, water will diffuse into the vacuole, placing pressure on the cell wall. If water loss leads to a significant decline in turgor pressure, the cell will plasmolyse.

Turgor pressure exerted by vacuoles is also helpful for cellular elongation: as the cell wall is partially degraded by the action of auxins, the less rigid wall is expanded by the pressure coming from within the vacuole.

Vacuoles can help some plant cells to reach considerable size. Another function of a central vacuole is that it pushes all contents of the cell's cytoplasm against the cellular membrane, and thus keeps the chloroplasts closer to light.

The vacuole also stores the pigments in flowers and fruits.

Animals:

Vacuoles in animals are a part of the processes of exocytosis and endocytosis. Exocytosis is the extrusion process of proteins from the Golgi apparatus initially enter secretory granules, where processing of prohormones to the mature hormones occurs before exocytosis, and also allows the animal cell to rid waste products. Endocytosis is the reverse of exocytosis. There are various types.

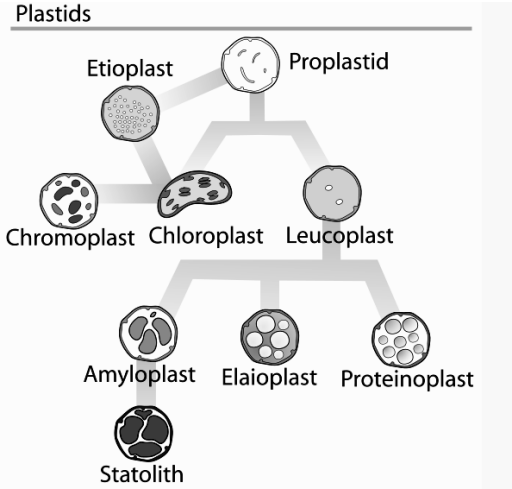
Phagocytosis ("cell eating") is the process by which bacteria, dead tissue, or other bits of material visible under the microscope are engulfed by cells. The material makes contact with the cell membrane, which then invaginates. The invagination is pinched off, leaving the engulfed material in the membrane-enclosed vacuole and the cell membrane intact.

Pinocytosis ("cell drinking") is essentially the same process, the difference being that the substances ingested are in solution and not visible under the microscope.

Hydropic (vacuolar) changes are of importance of identifying various pathologies, such as the reversible cell swelling in renal tubules caused by hypoperfusion of the kidneys during open heart surgery.

B7. Plastids

Plastids are major organelles found in plants and algae. Plastids often contain pigments used in photosynthesis, and the types of pigments present can change or determine the cell's color. Plastids are also the site of manufacture and storage of important chemical compounds used by the cell.



Plastids in plants: Plastids are responsible for photosynthesis, storage of products like starch and for the synthesis of many classes of molecules such as fatty acids and terpenes which are needed as cellular building blocks and/or for the function of the plant. Depending on their morphology and function, plastids have the ability to differentiate, or redifferentiate, between these and other forms. All plastids are derived from proplastids (formerly "eoplasts", eo-: dawn, early), which are present in the meristematic regions of the plant. Proplastids and young chloroplasts commonly divide, but more mature chloroplasts also have this capacity.

In plants, plastids may differentiate into several forms, depending upon which function they need to play in the cell. Undifferentiated plastids (proplastids) may develop into any of the following plastids:

- **Chloroplasts:** for photosynthesis; see also etioplasts, the predecessors of chloroplasts
- **Chromoplasts:** for pigment synthesis and storage
- **Leucoplasts**: for monoterpene synthesis; leucoplasts sometimes differentiate into more specialized plastids:
- **Amyloplasts:** for starch storage
- **Statoliths**: for detecting gravity
- **Elaioplasts:** for storing fat
- **Proteinoplasts**: for storing and modifying protein

Each plastid creates multiple copies of the circular 75-250 kilo bases plastid genome. The number of genome copies per plastid is flexible, ranging from more than 1000 in rapidly dividing cells, which generally contain few plastids, to 100 or fewer in mature cells, where plastid divisions has given rise to a large number of plastids. The plastid genome contains about 100 genes encoding ribosomal and transfer ribonucleic acids (rRNAs and tRNAs) as well as proteins involved in photosynthesis and plastid gene transcription and translation. However, these proteins only represent a small fraction of the total protein set-up necessary to build and maintain the structure and function of a particular type of plastid. Nuclear genes encode the vast majority of plastid proteins, and the expression of plastid genes and nuclear genes is tightly co-regulated to allow proper development of plastids in relation to cell differentiation.

Plastid DNA exists as large protein-DNA complexes associated with the inner envelope membrane and called 'plastid nucleoids'. Each nucleoid particle may contain more than 10 copies of the plastid DNA. The proplastid contains a single nucleoid located in the centre of the plastid. The developing plastid has many nucleoids, localized at the periphery of the plastid, bound to the inner envelope membrane. During the development of proplastids to chloroplasts, and when plastids convert from one type to another, nucleoids change in morphology, size and location within the organelle. The remodelling of nucleoids is believed to occur by modifications to the composition and abundance of nucleoid proteins.

Many plastids, particularly those responsible for photosynthesis, possess numerous internal membrane layers.

In plant cells long thin protuberances called stromules sometimes form and extend from the main plastid body into the cytosol and interconnect several plastids. Proteins, and presumably smaller molecules, can move within stromules. Most cultured cells that are relatively large compared to other plant cells have very long and abundant stromules that extend to the cell periphery.

Plastids in algae: In algae, the term leucoplast (leukoplast) is used for all unpigmented plastids. Their function differ from the leukoplasts in plants. Etioplast, amyloplast and chromoplast are plant-specific and do not occur in algae. Algal plastids may also differ from plant plastids in that they contain pyrenoids.

Inheritance of plastids: Most plants inherit the plastids from only one parent. Angiosperms generally inherit plastids from the mother, while many gymnosperms inherit plastids from the father. Algae also inherit plastids from only one parent. The plastid DNA of the other parent is thus completely lost.

In normal intraspecific crossings (resulting in normal hybrids of one species), the inheritance of plastid DNA appears to be quite strictly 100% uniparental. In interspecific hybridisations, however, the inheritance of plastids appears to be more erratic. Although plastids inherit mainly maternally in interspecific hybridisations, there are many reports of hybrids of flowering plants that contain plastids of the father.

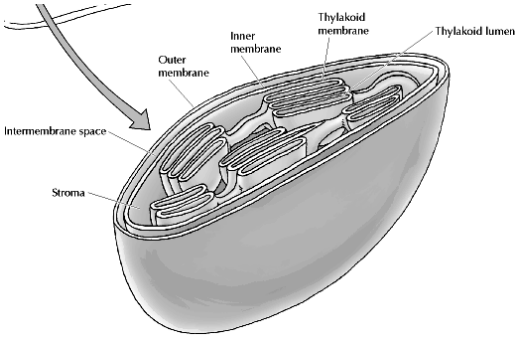
Origin of plastids: Plastids are thought to have originated from endosymbiotic cyanobacteria. They developed around 1500 mya and allowed eukaryotes to carry out oxygenic photosynthesis. Due to a split-up into three evolutionary lineages, the plastids are named differently: chloroplasts in green algae and plants, rhodoplasts in red algae and cyanelles in the glaucophytes. The plastids differ by their pigmentation, but also in ultrastructure. The chloroplasts e.g. have lost all phycobilisomes, the light harvesting complexes found in cyanobacteria, red algae and glaucophytes, but - only in plants and in closely related green algae - contain stroma and grana thylakoids. The glaucocystophycean plastid - in contrast to the chloroplasts and the rhodoplasts - is still surrounded by a remains of the

cyanobacterial cell wall. All these primary plastids are surrounded by two membranes.
Complex plastids start by secondary endosymbiosis, when a eukaryote engulfs a red or green alga and retains the algal plastid, which is typically surrounded by more than two membranes. In some cases these plastids may be reduced in their metabolic and/or photosynthetic capacity. Algae with complex plastids derived by secondary endosymbiosis of a red alga include the heterokonts, haptophytes, cryptomonads, and most dinoflagellates (= rhodoplasts). Those that endosymbiosed a green alga include the euglenids and chlorarachniophytes (= chloroplasts). The Apicomplexa, a phylum of obligate parasitic protozoa including the causative agents of malaria (Plasmodium spp.), toxoplasmosis (Toxoplasma gondii), and many other human or animal diseases also harbor a complex plastid (although this organelle has been lost in some apicomplexans, such as Cryptosporidium parvum, which causes cryptosporidiosis). The 'apicoplast' is no longer capable of photosynthesis, but is an essential organelle, and a promising target for antiparasitic drug development.

Some dinoflagellates take up algae as food and keep the plastid of the digested alga to profit from the photosynthesis; after a while the plastids are also digested. These captured plastids are known as kleptoplastids.

B8. Chloroplast

Chloroplasts, the organelles responsible for photosynthesis, are in many respects similar to mitochondria. Both chloroplasts and mitochondria function to generate metabolic energy, evolved by endosymbiosis, contain their own genetic systems, and replicate by division. However, chloroplasts are larger and more complex than mitochondria, and they perform several critical tasks in addition to the generation of ATP. Most importantly, chloroplasts are responsible for the photosynthetic conversion of CO_2 to carbohydrates. In addition, chloroplasts

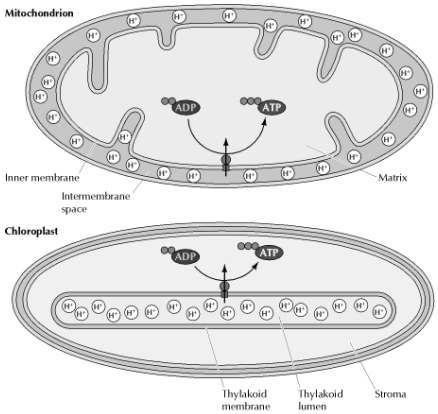


synthesize amino acids, fatty acids, and the lipid components of their own membranes. The reduction of nitrite (NO_2^-) to ammonia (NH_3), an essential step in the incorporation of nitrogen into organic compounds, also occurs in chloroplasts. Moreover, chloroplasts are only one of several types of related organelles (plastids) that play a variety of roles in plant cells.

The Structure and Function of Chloroplasts

Plant chloroplasts are large organelles (5 to 10 mm long) that, like mitochondria, are bounded by a double membrane called the chloroplast envelope (Figure). In addition to the inner and outer membranes of the envelope, chloroplasts have a third internal membrane system, called the thylakoid membrane. The thylakoid membrane forms a network of flattened discs called thylakoids, which are frequently arranged in stacks called grana. Because of this three-membrane structure, the internal organization of chloroplasts is more complex than that of mitochondria. In particular, their three membranes divide chloroplasts into three distinct internal compartments: (1) the intermembrane space between the two membranes of the chloroplast envelope; (2) the stroma, which lies inside the envelope but outside the thylakoid membrane; and (3) the thylakoid lumen.

Despite this greater complexity, the membranes of chloroplasts have clear functional similarities with those of mitochondria --as expected, given the role of both organelles in the chemiosmotic generation of ATP. The outer membrane of the chloroplast envelope, like that of mitochondria, contains porins and is therefore freely permeable to small molecules. In contrast, the inner membrane is impermeable to ions and metabolites, which are therefore able to enter chloroplasts only via specific membrane transporters. These properties of the inner and outer membranes of the chloroplast envelope are similar to the inner and outer membranes of mitochondria: In both cases the inner membrane restricts the passage of molecules between the cytosol and the interior of the organelle. The chloroplast stroma is also equivalent in function to the mitochondrial matrix: It contains the chloroplast genetic system and a variety of metabolic enzymes, including those responsible for the critical conversion of CO_2 to carbohydrates during photosynthesis.



The major difference between chloroplasts and mitochondria, in terms of both structure and function, is the thylakoid membrane. This membrane is of central importance in chloroplasts, where it fills the role of the inner mitochondrial membrane in electron transport and the chemiosmotic generation of ATP (Fig). The inner membrane of the chloroplast envelope (which is not folded into cristae) does not function in photosynthesis. Instead, the chloroplast electron transport system is located in the thylakoid membrane, and protons are pumped across this membrane from the stroma to the thylakoid lumen. The resulting electrochemical gradient then drives ATP synthesis as protons cross back into the stroma. In terms of its role in generation of metabolic energy, the thylakoid membrane of chloroplasts is thus equivalent to the inner membrane of mitochondria.

The Chloroplast Genome

Like mitochondria, chloroplasts contain their own genetic system, reflecting their evolutionary origins from photosynthetic bacteria. The genomes of chloroplasts are similar to those of mitochondria in that they consist of circular DNA molecules present in multiple copies per organelle. However, chloroplast genomes are larger and

more complex than those of mitochondria, ranging from 120 to 160 kb and containing approximately 120 genes.

Both the ribosomal and transfer RNAs used for translation of chloroplast mRNAs are encoded by the organelle genome. These include four rRNAs (23S, 16S, 5S, and 4.5S) and 30 tRNA species. In contrast to the smaller number of tRNAs encoded by the mitochondrial genome, the chloroplast tRNAs are sufficient to translate all the mRNA codons according to the universal genetic code. In addition to these RNA components of the translation system, the chloroplast genome encodes about 20 ribosomal proteins, which represent approximately a third of the proteins of chloroplast ribosomes. Some subunits of RNA polymerase are also encoded by chloroplasts, although additional RNA polymerase subunits and other factors needed for chloroplast gene expression are encoded in the nucleus.
The chloroplast genome also encodes approximately 30 proteins that are involved in photosynthesis, including components of photosystems I and II, of the cytochrome bf complex, and of ATP synthase. In addition, one of the subunits of ribulose bisphosphate carboxylase (rubisco) is encoded by chloroplast DNA. Rubisco is the critical enzyme that catalyzes the addition of CO_2 to ribulose-1,5-bisphosphate during the Calvin cycle. Not only is it the major protein component of the chloroplast stroma, but it is also thought to be the single most abundant protein on Earth, so it is noteworthy that one of its subunits is encoded by the chloroplast genome.

B9. Structure and Function of Cytoskeleton and its role in motility:

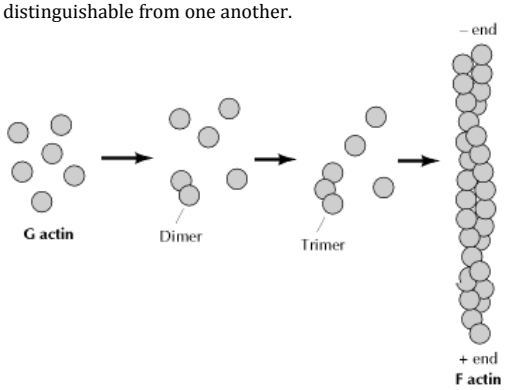
The membrane-enclosed organelles discussed in the preceding section constitute one level of the organizational substructure of eukaryotic cells. A further level of organization is provided by the cytoskeleton, which consists of a network of protein filaments extending throughout the cytoplasm of all eukaryotic cells. The cytoskeleton provides a structural framework for the cell, serving as a scaffold that determines cell shape and the general organization of the cytoplasm. In addition to playing this structural role, the cytoskeleton is responsible for cell movements. These include not only the movements of entire cells, but also the internal transport of organelles and other structures (such as mitotic chromosomes) through the cytoplasm. Importantly, the cytoskeleton is much less rigid and permanent than its name implies. Rather, it is a dynamic structure that is continually reorganized as cells move and change shape, for example, during cell division.

The cytoskeleton is composed of three principal types of protein filaments: **actin filaments, intermediate filaments, and microtubules**, which are held together and linked to subcellular organelles and the plasma membrane by a variety of accessory proteins. This chapter discusses the structure and organization of each of these three major components of the cytoskeleton, as well as their roles in cell motility, organelle transport, cell division, and other types of cell movements.

Structure and Organization of Actin Filaments

The major cytoskeletal protein of most cells is actin, which polymerizes to form actin filaments—thin, flexible fibers approximately 7 nm in diameter and up to several micrometers in length. Within the cell, actin filaments (also called microfilaments) are organized into higher-order structures, forming bundles or three-dimensional networks with the properties of semisolid gels. The assembly and disassembly of actin filaments, their crosslinking into bundles and networks, and their association with other cell structures (such as the plasma membrane) are regulated by a variety of actin-binding proteins, which are critical components of the actin cytoskeleton. Actin filaments are particularly abundant beneath the plasma membrane, where they form a network that provides mechanical support, determines cell shape, and allows movement of the cell surface, thereby enabling cells to migrate, engulf particles, and divide.

Individual actin molecules are globular proteins of 375 amino acids (43 kd). Each actin monomer (globular [G] actin) has tight binding sites that mediate head-to-tail interactions with two other actin monomers, so actin monomers polymerize to form filaments (filamentous [F] actin). Each monomer is rotated by 166° in the filaments, which therefore have the appearance of a double-stranded helix. Because all the actin monomers are oriented in the same direction, actin filaments have a distinct polarity and their ends (called the plus and minus ends) are distinguishable from one another.

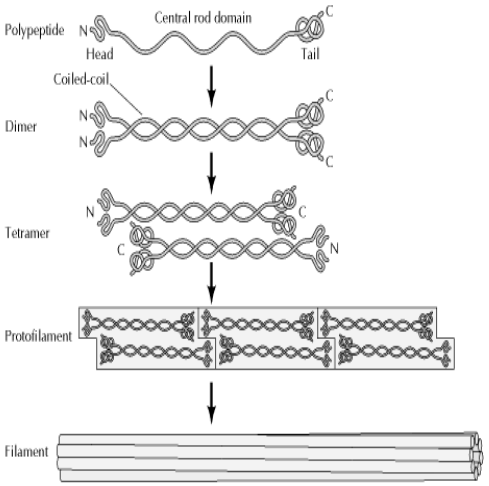


It is noteworthy that several drugs useful in cell biology act by binding to actin and affecting its polymerization. For example, the cytochalasins bind to the plus ends of actin filaments and block their elongation. This results in changes in cell shape as well as inhibition of some types of cell movements (e.g., cell division following mitosis), indicating that actin polymerization is required for these processes. Another drug, phalloidin, binds tightly to actin filaments and prevents their dissociation into individual actin molecules. Phalloidin labeled with a fluorescent dye is frequently used to visualize actin filaments by fluorescence microscopy.

Intermediate Filaments

Intermediate filaments have a diameter of about 10 nm, which is intermediate between the diameters of the two other principal elements of the cytoskeleton, actin filaments (about 7 nm) and microtubules (about 25 nm). In contrast to actin filaments and microtubules, the intermediate filaments are not directly involved in cell movements. Instead, they appear to play basically a structural role by providing mechanical strength to cells and tissues.

Intermediate Filament Proteins: Whereas actin filaments and microtubules are polymers of single types of proteins (actin and tubulin, respectively), intermediate filaments are composed of a variety of proteins that are expressed in different types of cells. More than 50 different intermediate filament proteins have been identified and classified into six groups based on similarities between their amino acid sequences.



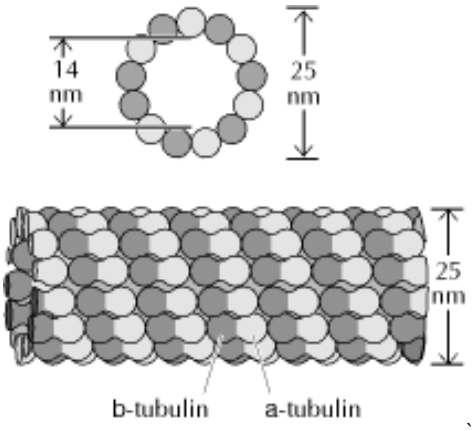
Types I and II consist of two groups of **keratins**, each consisting of about 15 different proteins, which are expressed in epithelial cells. Some type I and II keratins (called hard keratins) are used for production of structures such as hair, nails, and horns. The type III intermediate filament proteins include **vimentin**, which is found in a variety of different kinds of cells, including fibroblasts, smooth muscle cells, and white blood cells. Another type III protein, **desmin**, is specifically expressed in muscle cells, where it connects the Z discs of individual contractile elements. The type IV intermediate filament proteins include the three **neurofilament (NF) proteins** (designated NF-L, NF-M, and NF-H for light, medium, and heavy, respectively). These proteins form the major intermediate filaments of many types of mature neurons. The type V intermediate filament proteins are the **nuclear lamins**, which are found in most eukaryotic cells. Rather than being part of the cytoskeleton, the nuclear lamins are components of the nuclear envelope. They also differ from the other intermediate filament proteins in that they assemble to form an orthogonal meshwork underlying the nuclear membrane. Despite considerable diversity in size and amino acid sequence, the various intermediate filament proteins share a common structural organization.

Microtubules

Microtubules, the third principal component of the cytoskeleton, are rigid hollow rods approximately 25 nm in diameter. Like actin filaments, microtubules are dynamic structures that undergo continual assembly and disassembly within the cell. They function both to determine cell shape and in a variety of cell movements, including some forms of cell locomotion, the intracellular transport of organelles, and the separation of chromosomes during mitosis.

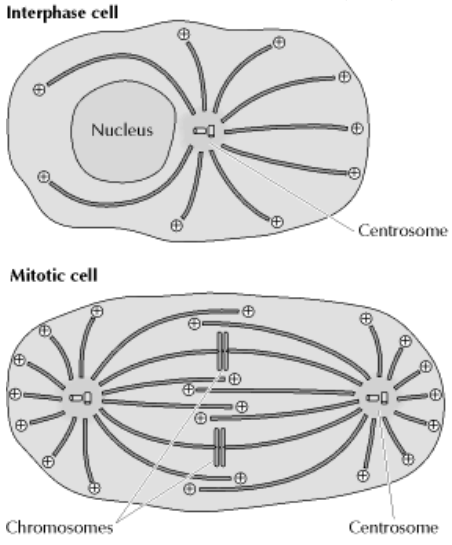
Structure, Assembly, and Dynamic Instability of Microtubules:

In contrast to intermediate filaments, which are composed of a variety of different fibrous proteins, microtubules are composed of a single type of globular protein, called **tubulin**. Tubulin is a dimer consisting of two closely related 55-kd polypeptides, α-tubulin and β-tubulin. Like actin, both α- and β-tubulin are encoded by small families of related genes. In addition, a third type of tubulin (γ-tubulin) is specifically localized to the centrosome, where it plays a critical role in initiating microtubule assembly.



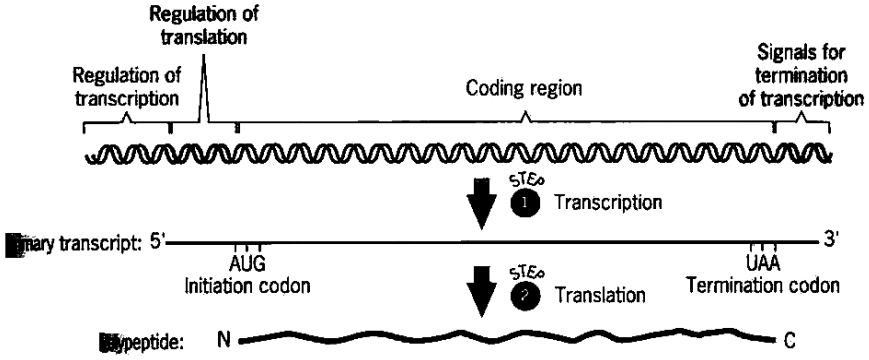
Tubulin dimers polymerize to form microtubules, which generally consist of 13 linear protofilaments assembled around a hollow core. The protofilaments, which are composed of head-to-tail arrays of tubulin dimers, are arranged in parallel. Consequently, microtubules (like actin filaments) are polar structures with two distinct ends: a fast-growing plus end and a slow-growing minus end. This polarity is an important consideration in determining the direction of movement along microtubules, just as the polarity of actin filaments defines the direction of myosin movement.

Tubulin dimers can depolymerize as well as polymerize, and microtubules can undergo rapid cycles of assembly and disassembly. Both α- and β-tubulin bind GTP, which functions analogously to the ATP bound to actin to regulate polymerization. In particular, the GTP bound to b-tubulin (though not that bound to a-tubulin) is hydrolyzed to GDP during or shortly after polymerization. Whether a microtubule grows or shrinks is determined by the rate of tubulin addition relative to the rate of GTP hydrolysis.

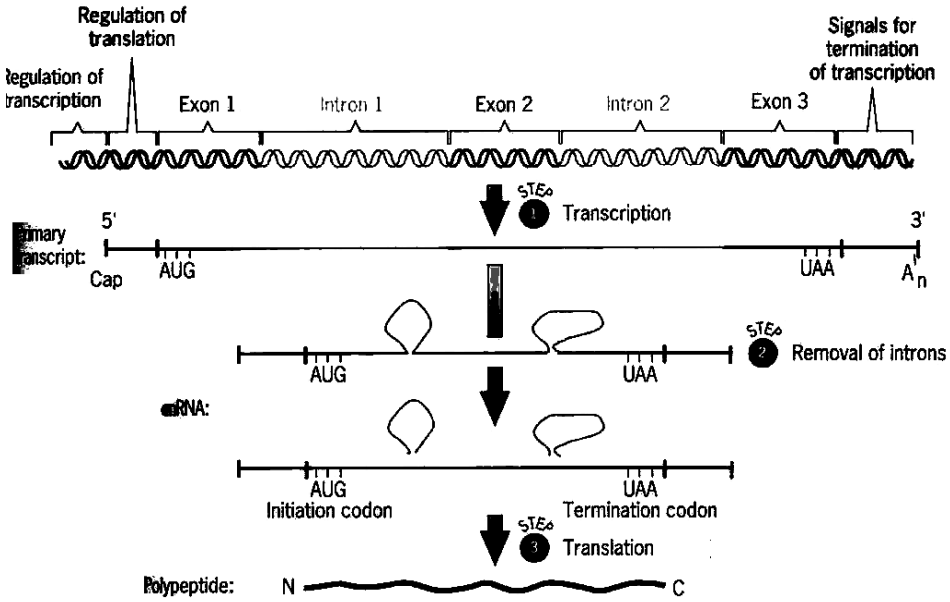


C1. Organization of Genes:

In molecular terms, a gene commonly is defined as *the entire nucleic acid sequence that is necessary for the synthesis of a functional polypeptide.* According to this definition, a gene includes more than the nucleotides encoding the amino acid sequence of a protein, referred to as the *coding region.* A gene also includes all the DNA sequences required for synthesis of a particular RNA transcript.



(a) A typical prokaryotic gene.



(b) A typical eukaryotic gene.

Figure Gene structure. Prokaryotic genes usually contain uninterrupted coding sequences (*a*) whereas the coding sequences of eukaryotic genes are commonly interrupted by noncoding sequences (*b*).

- Thus, a gene is the entire DNA sequence required for synthesis of a functional protein or RNA molecule. In addition to the coding regions (exons), a gene includes transcription-control regions and sometimes introns. Although the majority of genes encode proteins, some encode tRNAs, rRNAs, and other types of RNA.
- Most bacterial genes have no introns, whereas most genes of multicellular organisms do. The introns in human genes encoding average-size proteins are often much longer than the exons.
- Many bacterial proteins with related functions are encoded by contiguous genes regulated by a single transcription-control region. This type of gene cluster, called an operon, is transcribed into a single, polycistronic mRNA, which is translated to yield several different proteins.
- Most eukaryotic transcription units are transcribed into monocistronic mRNAs, each of which is translated into a single protein.
- The primary transcript produced from a simple eukaryotic transcription unit is processed into a single type of mRNA. The primary transcript produced from a complex eukaryotic transcription unit can be processed into two or more different mRNAs depending on the choice of splice sites and/or polyadenylation site. In the case of many complex units, one mRNA is produced in one cell type, while an alternative mRNA is produced in a different cell type.

Organization of Genes and Non-coding DNA

- The genomes of prokaryotes and lower eukaryotes contain few nonfunctional sequences, whereas vertebrate genomes contain many sequences that do not code for RNAs or have any structural or regulatory function. Only about 5 percent of the genomic DNA in humans encodes proteins or functional RNAs.
- The lack of a consistent relationship between the amount of DNA in the haploid chromosomes of an animal or plant and its phylogenetic complexity is called the **C-value paradox**.

The Human Genome: This term is used to describe the different types of sequence that together make up the DNA in a human cell. The DNA in the human genome is about 3 billion base pairs (bp) long and is estimated to contain 30000 genes. The DNA is arranged as a set of 23 chromosomes each of which is a single, double-stranded DNA molecule 55 - 250 million bp long. The genes and gene-related sequences account for about 25% of the DNA *(Fig.*). The remainder is called extragenic DNA and has no known function.

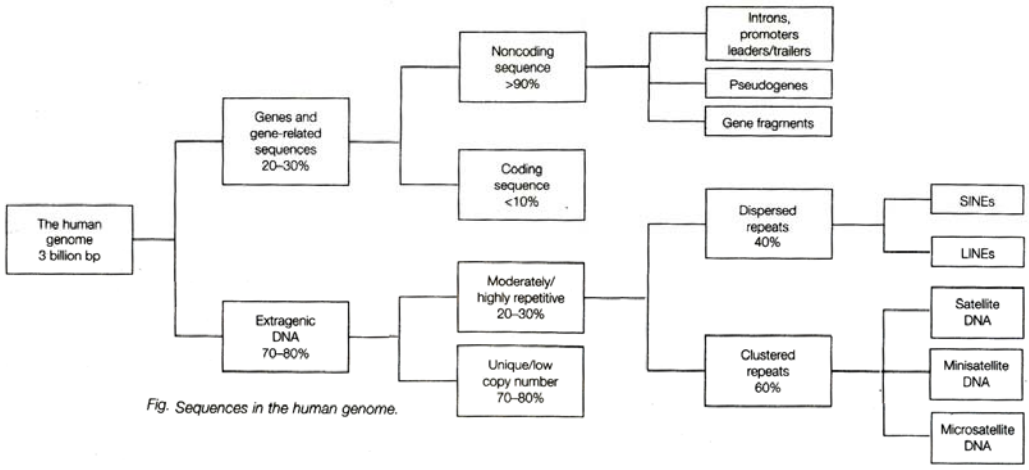


Fig. Sequences in the human genome.

Genes and Gene-related sequences (20-30%):

Genes: The coding information in a gene is present as a series of segments of DNA sequence called **exons** separated from each other by intervening non-coding sequences called **introns**. Genes vary greatly in size and also with respect to the number and sizes of the introns. Some genes such as the histone H4 gene are just a few hundred base pairs long. Others, such as the Factor VIII gene, are several hundred kilobase pairs (kbp) in length and contain many large introns such that the actual coding sequence accounts for just a few percent of the total gene sequence. Additional sequences are present which are associated with genes. Leader and trailer sequences occur at the 5' and 3' ends of the gene which are transcribed but not translated. Promoter sequences occur upstream of the point where transcription begins and regulate synthesis of mRNA from the gene. The promoter may extend up to about 1 kbp upstream but other regulatory sequences that influence transcription may occur at sites much further away.

Gene families: Some genes exist as a number of copies with identical or related sequences that can be grouped into families. Gene families may be organized in a number of ways:

(i) all of the genes in the family occur at the same chromosomal locus. An example of this is the growth hormone gene family whose five members are clustered on chromosome 17;
(ii) the genes belonging to the family may occur at different loci. For example; the five members of the aldolase gene family are on different chromosomes;
(iii) the genes of a family may exist as a series of clusters on different chromosomes. An example is the homeobox genes which occur as four clusters on separate chromosomes each containing about 10 individual genes.

In some multigene families all the genes are identical and may encode a protein required in large amounts by the cell such as histones. In other families, the genes are not identical but show some sequence divergence. In some cases the divergence is so great that the genes encode proteins that are related but have distinctive properties. An example of this is the α- and β-globin gene-families whose members are expressed at different stages of embryonic development and in adults.

Psuedogenes: These are diverged members of gene families that have acquired one or more inactivating mutations so that they are no longer able to function and do not produce biologically active protein. A pseudogene is simply a mutated version of a parent gene. Often the mutations present are nonsense mutations which generate stop codons and result in premature termination of translation. Related to pseudogenes are processed pseudogenes which are DNA copies of mRNA. **Processed pseudogenes** lack introns and because they do not have a promoter they are not transcribed and do not result in protein production. It is not clear how processed pseudogenes arose in the, human genome but they may have been derived by reverse transcription of mRNA into double-stranded DNA which was then inserted into the genome. Another group of inactive genes are **gene fragments** which lack the 5' or 3' region of the parent gene. These are thought to have arisen by a deletion event or by recombination that split the parent gene.

Extragenic DNA(70-80%)

This part of the human genome is composed of sequences that exist in addition to the genes and gene-related sequences described above *(Fig.*). Extragenic DNA is composed of sequences that are not part of a gene (exons and introns), not associated with a gene (leader and trailer sequences, promoters and distant regulatory elements) and not a pseudogene or a gene fragment. Although extragenic sequences account for most of the DNA in the human genome (70-80%), they have no known function. Most of the extragenic DNA sequences (70-80%) are unique or exist as a small number of copies. The remainder (20-30%) are moderately or highly repeated sequences that may be dispersed throughout the genome or lined up end on end as long tandem arrays.

Dispersed Repetitive Sequences: Two types of dispersed repetitive sequence exist known as short and long inter-spersed nuclear elements, abbreviated to **SINEs and LINEs**.

- **SINEs:** The best known example of SINEs is **Alu elements**. These sequences are not identical but they are similar enough to be classed as a family. They have an average length of 250 bp and almost 1 million copies exist which are very widely dispersed throughout the genome occurring in most places including the introns of some genes. The origin of Alu elements is uncertain but they are believed to have arisen as one or more processed pseudogenes. It is suggested that at one stage these sequences acquired transposable activity which allowed them to replicate and move to different sites in the genome.
- **LINEs:** They are similar to SINEs but have longer sequences. A well known example is the **L1 LINE** which is 6500 bp long and is present as 60,000 copies. LINEs are a type of **retroelement**. These are sequences that are capable of moving through the genome by a process called transposition which allows them to copy themselves by reverse transcription and to insert the copy into the genome at a distant site (Fig). Most L1 elements are truncated, but the full length version has all the components needed for transposition including a gene that could code for reverse transcriptase.

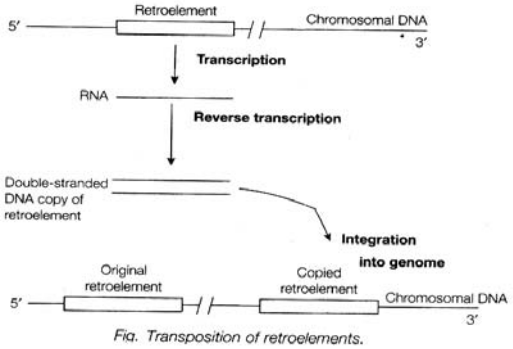


Fig. Transposition of retroelements.

Clustered Repetitive Sequences: The human genome contains extensive regions in which repetitive sequences are arranged end on end as long tandem arrays. This is called **satellite DNA** and occurs in various forms, each with a different repeat unit. because they are separated from the bulk of cellular DNA by equilibrium density-gradient centrifugation. However, not all simple-sequence DNAs separate from the bulk of cellular DNA during centrifugation Simple-sequence DNA, which consists largely of very short sequences repeated in long tandem arrays, is preferentially located in centromeres, telomeres, and specific locations within the arms of particular chromosomes where it may have a structural role. Three classes of satellite DNA have been defined based on size of the cluster and the length of the repeat unit:

- **Satellite DNA**: Satellite DNA was the first type to be identified in the human genome and contains repeat units of up to 200 bp arranged in clusters between 100 and 5000 kbp;
- **Minisatellite DNA**: Minisatellite DNA occurs in smaller clusters up to 20 kbp in length and has repeat units up to 25 bp; The length of a particular simple-sequence tandem array is quite variable among individuals in a species, probably because of unequal crossing over during meiosis. Differences in the lengths of mini satellite DNA forms the basis for DNA fingerprinting
- **Microsatellite DNA:** Microsatellite DNA occurs in short clusters, usually less than 150 bp, and has a repeat unit which is usually 4 bp or less. Microsatellite DNA is very common. One type, which contains the dinucleotide CA as the repeat sequence, accounts for 0.5% of the entire genome. Mononucleotide repeats, consisting of a single repeated base, account for a further 0.3%.

Variable number of tandem repeats: The length of Microsatellite DNA sequences such as the CA repeats and Mini satellite DNA varies from person to person. At any given location in the genome, the number of repeat units present in an individual microsatellite DNA sequence may vary by as many as 10 or more. As such, these sequences are known as variable number tandem repeats or VNTRs. The polymerase chain reaction (PCR) can be used to analyse VNTRs *(Fig.*). Using primers specific for unique sequences on either side of the repeat sequence, DNA molecules of varying lengths are amplified depending on how many repeats are present. By examining several VNTRs it is possible to build up a genetic profile that is unique to the individual being tested.

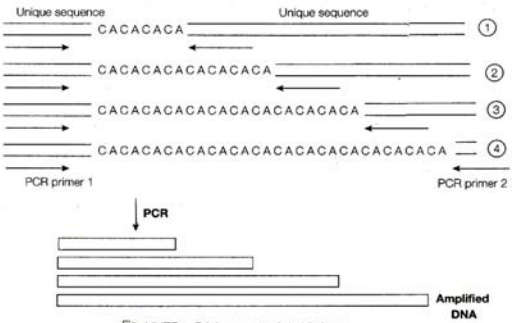


Fig. VNTRs. CA is repeated 4–16 times. PCR using primers specific for flanking unique sequence produces amplified DNA of varying lengths.

VNTRs have a number of very useful applications based on establishing the identity of an individual. They can be used in forensic science to link suspects to a crime using biological material recovered from the scene. They also have a number of uses in medicine including matching patients undergoing transplants with donor organs, identifying carriers of genetic diseases and establishing paternity. VNTRs have also been widely used in the human genome mapping project as gene markers to identify the location and order of genes on chromosomes.

C2. GENOME ORGANIZATION

Bacterial Chromosomes:

In most bacterial cells, genes are encoded on large circular chromosomes. The circular nature of bacterial chromosomes was first discovered by analyzing the frequency of genetic recombination between mutant genes that produced easily assayed phenotypes, such as the inability to grow in the absence of a specific amino acid or the inability to grow on a particular sugar. Eventually, the circular structure of the *E. coli* chromosome was observed directly in autoradiographs of DNA molecules from cells grown in the presence of ^{3}Hlabeled thymine, which is incorporated only into DNA. Such studies showed that the *E. coli* chromosome has a total length of 1 mm. The *E. coli* chromosomal DNA molecule replicates from a single replication origin, similar to the small circular plasmid DNA molecule.

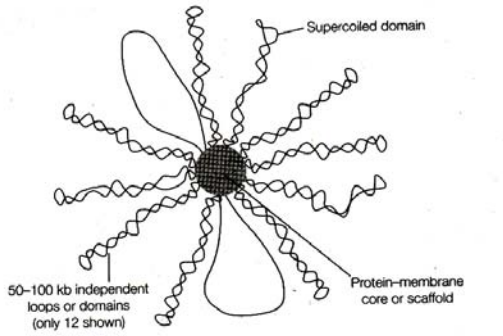


Fig. A schematic view of the structure of the E. coli *chromosome (4600 kb) as visualized by electron microscopy. The thin line is the DNA double helix.*

The 1-mm-long DNA molecule of the *E. coli* chromosome is contained within cells that are only about 2 μm long and about 0.5 1 μm wide. A free DNA molecule of this size would form a random coil about 1000 times the volume of an *E. coli* cell. However, several mechanisms operate to compact *E. coli* chromosomal DNA sufficiently to fit inside the bacterial cell. For example, the large volume filled by free DNA is due largely to charge repulsion between the negatively charged phosphate groups. In the cell, this effect is counteracted by association of the DNA with positively charged polyamines, such as spermine and spermidine, which shield the negative charges of the DNA phosphate groups:

In addition, numerous small protein molecules associate with the chromosomal DNA, causing it to fold into a more compact structure. The most abundant of these proteins, H-NS, is a dimer of a 15.6-kDa polypeptide. H-NS binds DNA tightly and compacts it considerably, as indicated by the increased sedimentation rate and decreased viscosity of DNA associated with H-NS compared with free DNA. There are about 20,000 H-NS molecules per *E. coli* cell, enough for one H-NS dimer per ≈400 base pairs of DNA.

Finally, *E. coli* chromosomal DNA is tightly supercoiled — that is, twisted upon itself like the circular SV40 DNA. An *E. coli* enzyme called DNA gyrase uses energy from ATP hydrolysis to wind supercoils into DNA. Supercoiling contributes to the compaction necessary to fit chromosomal DNA into the bacterial cell. If all the supercoils were relaxed and the DNA spread out, it would appear as a single, replicating, circular DNA molecule.

Eukaryotic Nuclear DNA Associates with Histone Proteins to Form Chromatin

The most abundant proteins associated with eukaryotic DNA are histones, a family of basic proteins present in all eukaryotic nuclei. The five major types of histone proteins —termed H1, H2A, H2B, H3, and H4 —are rich in positively charged basic amino acids, which interact with the negatively charged phosphate groups in DNA. In a fraction of the histone proteins of most cells, some of the basic amino acid side chains are modified by post-translational addition of acetyl (CH_3COO^-), phosphate, or methyl groups, neutralizing the positive charge of the side chain or converting it to a negative charge. The amino acid sequences of four histones (H2A, H2B, H3, and H4) are remarkably similar among distantly related species. For example, the sequences of histone H3 from sea urchin tissue and of H3 from calf thymus are identical except for a single amino acid, and only four amino acids are different in H3 from the garden pea and that from calf thymus. Minor histone variants encoded by genes that differ from the highly conserved major types also exist, particularly in vertebrates. The amino acid sequence of H1 varies more from organism to organism than do the sequences of the other major histones. In certain tissues, H1 is replaced by special histones. For example, in the nucleated red blood cells of birds, a histone termed H5 is present in place of H1. The similarity in sequence among histones from all eukaryotes indicates that they fold into very similar three-dimensional conformations, which were optimized for histone function early in evolution in a common ancestor of all modern eukaryotes.

Chromatin: The total length of DNA in eukaryotic cell depends on the species but it can be thousands of times as much as in a prokaryotic genome, and is made up of a number of discrete bodies called chromosomes (46 in humans). The DNA in each chromosome is believed to be a single linear molecule, which can be up to several centimeters long. This entire DNA must be packaged into the nucleus, a space of approximately the same volume as a bacterial cell; in fact, in their most highly condensed forms, the chromosomes have an enormously high DNA concentration of perhaps 200 mg ml-1. This feat of packing is accomplished by the formation of a highly organized complex of DNA and protein, known as chromatin, a nucleoprotein complex. More than 50% of the mass of chromatin is protein. Chromosomes greatly alter their level of compactness as cells progress through the cell cycle, varying between highly condensed chromosomes at metaphase (just before cell division), and very much more diffuse structures in interphase. This implies the existence of different levels of organization of chromatin.

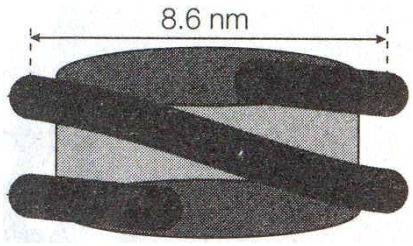
Histones: Most of the protein in eukaryotic chromatin consists of histones, of which there are five families, or classes: H2A, H2B, H3 and H4, known as the core histones, and H1. The core his tones are small proteins, with masses between 10 and 20 kDa, and H1 histones are a little larger at around 23 kDa. All histone proteins have a large positive charge; between 20 and 30% of their sequences consist of the basic amino acids, lysine and arginine. This means that his tones will bind very strongly to the negatively charged DNA in forming chromatin.

Members of the same histone class are very highly conserved between relatively unrelated species, for example between plants and animals, which testifies to their crucial role in chromatin. Within a given species, there are normally a number of closely similar variants of a particular class, which may be expressed in different tissues, and at different stages in development. There is not much similarity in sequence between the different histone classes, but structural studies have shown that the classes do share a similar tertiary structure, suggesting that all histones are ultimately evolutionarily related.

H1 his tones are somewhat distinct from the other histone classes in a number of ways; in addition to their larger size, there is more variation in H1 sequences both between and within species than in the other classes. Histone H1 is more easily extracted from bulk chromatin, and seems to be present in roughly half the quantity of the other classes, of which there are very similar amounts. These facts suggest a specific and distinct role for histone H1 in chromatin structure.

Nucleosomes:

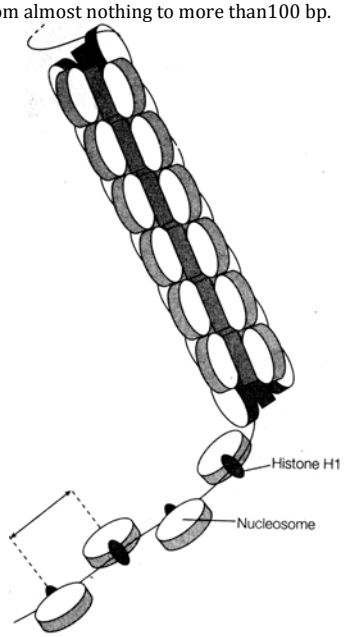
A number of studies in the I970s pointed to the existence of a basic unit of chromatin structure. Nucleases are enzymes which hydrolyze the phosphodiester bonds of nucleic acids. Exonucleases release single nucleotides from the ends of nucleic acid strands, whereas endonucleases cleave internal phosphodiester bonds. Treatment of chromatin with micrococcal nuclease, an endonuclease which cleaves double-stranded DNA, led to the isolation of DNA fragments with discrete sizes in multiples of approximately 200 bp. It was discovered that each 200 bp fragment is associated with an octamer Core of histone proteins, $(H2A)_2(H2B)_2(H3)_2(H4)_2$ which is why these are designated the core histones, and more loosely with one molecule of H1. The proteins protect the DNA from the action of micrococcal nuclease. More prolonged digestion with nuclease leads to the loss of Hl and yields a very resistant structure consisting of 146 bp of DNA associated very tightly with the histone octamer. This structure is known as the nucleosome core, and is structurally very similar whatever the source of the chromatin.



The structure of the nucleosome core particle is now known in considerable detail, from structural studies culminating in X-ray crystallography. The histone octamer forms a wedge-shaped disk, around which the 146 bp of DNA is wrapped in 1.8 turns in a left-handed direction. Figure 1 shows the basic features and the dimensions of the structure. The left-handed wrapping of the DNA around the nucleosome corresponds to negative supercoiling, that is the turns are superhelical turns. Although eukaryotic DNA is negatively supercoiled to a similar level as that of prokaryotes, on average, virtually all the supercoiling is accounted for by wrapping in nucleosomes and there is no unconstrained supercoiling.

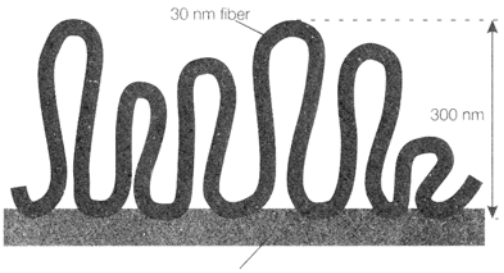
The Role of H1: One molecule of histone HI binds to the nucleosome, and acts to stabilize the point at which the DNA enters and leaves the nucleosome core. In the presence of HI, a further 20 bp of DNA is protected from nuclease digestion, making 166 bp in all, corresponding to two full turns around the histone octamer. A nucleosome core plus HI is known as a chromatosome. The larger size of HI compared with the core histones is due to the presence of an additional C-terminal tail, which serves to stabilize the DNA between the nucleosome cores. As stated above, H1 is more variable in sequence than the other histones and, in some cell types, it may be replaced by an extreme variant called histone H5, which binds chromatin particularly tightly, and is associated with DNA which is not undergoing transcription.

Linker DNA: In electron micrographs of nucleosome core particles on DNA under certain conditions, an array, sometimes called the 'beads on a string' structure is visible This comprises globular particles (nucleosomes), connected by thin strands of DNA. This linker DNA is the additional DNA required to make up the 200 bp nucleosomal repeat apparent in the micrococcal nuclease experiments (see above). The average length of linker DNA between core particles is 55 bp, but the length varies between species and tissues from almost nothing to more than100 bp.



The Solenoid-30 nm fiber: The presence of histone HI increases the organization of the 'beads on a string' to show a zig - zag structure in electron micrographs. With a change in the salt concentration, further organization of the nucleosomes into a fiber of 30 nm diameter takes place. Detailed studies of this process have suggested that the nucleosomes are wound into a higher order left-handed helix, dubbed a solenoid, with around six nucleosomes per turn. However, there is still some conjecture about the precise organization of the fiber structure, including the path of the linker DNA and the way in which different linker lengths might be incorporated into what seems to be a very uniform structure. Most chromosomal DNA *in vivo* is packaged into the 30 nm fiber.

Higher Order Structure: The organization of chromatin at the highest level seems rather similar to that of prokaryotic DNA. Electron micrographs of chromosomes which have been stripped of their histone proteins show a looped domain structure. Even the size of the loops is approximately the same, up to around 100 kb of DNA, although there are many more loops in a eukaryotic chromosome. The loops are constrained by interaction with a protein complex known as the nuclear matrix. The DNA in the loops is in the form of 30 nm fiber, and the loops form an array about 300 nm across.



C3. Transposons.

Mobile DNA

The second type of repetitious DNA in eukaryotic genomes, termed *interspersed repeats* (also known as *moderately repeated DNA,* or *intermediate-repeat DNA*) is composed of a very large number of copies of relatively few sequence families. These sequences, which are interspersed throughout mammalian genomes, make up ≈25–50 percent of mammalian DNA (≈45 percent of human DNA).

Because moderately repeated DNA sequences have the unique ability to "move" in the genome, they are called **mobile DNA elements** (or *transposable elements*). Although mobile DNA elements, ranging from hundreds to a few thousand base pairs in length, originally were discovered in eukaryotes, they also are found in prokaryotes. The process by which these sequences are copied and inserted into a new site in the genome is called **transposition.** Mobile DNA elements (or simply *mobile elements*) are essentially molecular symbiots that in most cases appear to have no specific function in the biology of their host organisms, but exist only to maintain themselves. For this reason, Francis Crick referred to them as "selfish DNA."

When transposition of eukaryotic mobile elements occurs in germ cells, the transposed sequences at their new sites can be passed on to succeeding generations. In this way, mobile elements have multiplied and slowly accumulated in eukaryotic genomes over evolutionary time. Since mobile elements are eliminated from eukaryotic genomes very slowly, they now constitute a significant portion of the genomes of many eukaryotes. Transposition also may occur within a somatic cell; in this case the transposed sequence is transmitted only to the daughter cells derived from that cell. In rare cases, this may lead to a somatic-cell mutation with detrimental phenotypic effects, for example, the inactivation of a tumor suppressor gene.

Movement of Mobile Elements Involves a DNA or an RNA Intermediate

Barbara McClintock discovered the first mobile elements while doing classical genetic experiments in maize (corn) during the 1940s. She characterized genetic entities that could move into and back out of genes, changing the phenotype of corn kernels. Her theories were very controversial until similar mobile elements were discovered in bacteria, where they were characterized as specific DNA sequences, and the molecular basis of their transposition was deciphered.

As research on mobile elements progressed, they were found to fall into two categories: (1) those that transpose directly as DNA and (2) those that transpose via an RNA intermediate transcribed from the mobile element by an RNA polymerase and then converted back into double-stranded DNA by a **reverse transcriptase** (Figure 1).

Mobile elements that transpose through a DNA intermediate are generally referred to as **DNA transposons.** Mobile elements that transpose to new sites in the genome via an RNA intermediate are called **retrotransposons** because their movement is analogous to the infectious process of retroviruses. Indeed, retroviruses can be thought of as retrotransposons that evolved genes encoding viral coats, thus allowing them to transpose between cells. Retrotransposons can be further classified on the basis of their specific mechanism of transposition.

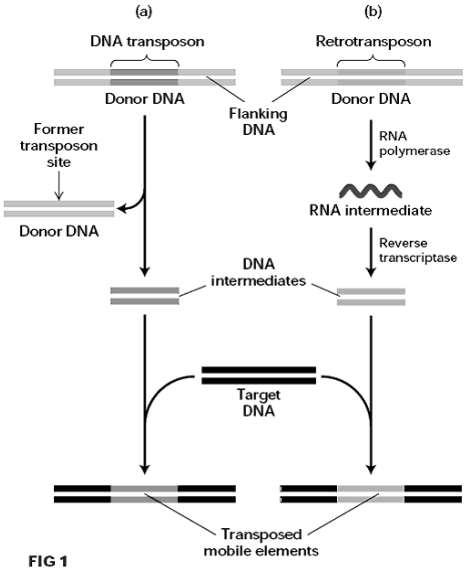


FIG 1
Classification of mobile elements into two major classes.

Mobile Elements That Move as DNA Are Present in Prokaryotes and Eukaryotes

Most mobile elements in bacteria transpose directly as DNA. In contrast, most mobile elements in eukaryotes are retrotransposons, but eukaryotic DNA transposons also occur. Indeed, the original mobile elements discovered by Barbara McClintock are DNA transposons.

Bacterial Insertion Sequences

The first molecular understanding of mobile elements came from the study of certain *E. coli* mutations caused by the spontaneous insertion of a DNA sequence, ≈1–2 kb long, into the middle of a gene. These inserted stretches of DNA are called *insertion sequences,* or *IS elements.* So far, more than 20 different IS elements have been found in *E. coli* and other bacteria.

Transposition of an IS element is a very rare event, occurring in only one in 10^5–10^7 cells per generation, depending on the IS element. Many transpositions inactivate essential genes, killing the host cell and the IS elements it carries. Therefore, higher rates of transposition would probably result in too great a mutation rate for the host cell to survive. However, since IS elements transpose more or less randomly, some transposed sequences enter nonessential regions of the genome (e.g., regions between genes), allowing the cell to survive. At a very low rate of transposition, most host cells survive and therefore propagate the symbiotic IS element. IS elements also can insert into plasmids or lysogenic viruses, and thus be transferred to other cells. When this happens, IS elements can transpose into the chromosomes of virgin cells. The general structure of IS elements is diagrammed in Figure 2.

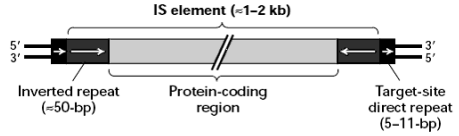


FIGURE 2 General structure of bacterial IS elements.

An *inverted repeat,* usually containing ≈50 base pairs, invariably is present at each end of an insertion sequence.

In an inverted repeat the 5 → 3sequence on one strand is repeated on the other strand, as:

5′ GAGC———GCTC 3′
3′ CTCG———CGAG 5′

Between the inverted repeats is a region that encodes *transposase,* an enzyme required for transposition of the IS element to a new site. The transposase is expressed at a very low rate, accounting for the very low frequency of transposition. An important hallmark of IS elements is the presence of a short *direct-repeat sequence,* containing 5–11 base pairs, immediately adjacent to both ends of the inserted element.

The *length* of the direct repeat is characteristic of each type of IS element, but its *sequence* depends on the target site where a particular copy of the IS element is inserted. When the sequence of a mutated gene containing an IS element is compared with the sequence of the wild-type gene before insertion, only one copy of the short direct-repeat sequence is found in the wild-type gene. Duplication of this target-site sequence to create the second direct repeat adjacent to an IS element occurs during the insertion process.

As depicted in Figure 3, transposition of an IS element is similar to a "cut-and-paste" operation in a word processing program. Transposase performs three functions in this process: it (1) precisely excises the IS element in the donor DNA, (2) makes staggered cuts in a short sequence in the target DNA, and (3) ligates the 3′ termini of the IS element to the 5′ ends of the cut donor DNA. Finally, a host-cell DNA polymerase fills in the single-stranded gaps, generating the short direct repeats that flank IS elements, and DNA ligase joins the free ends.

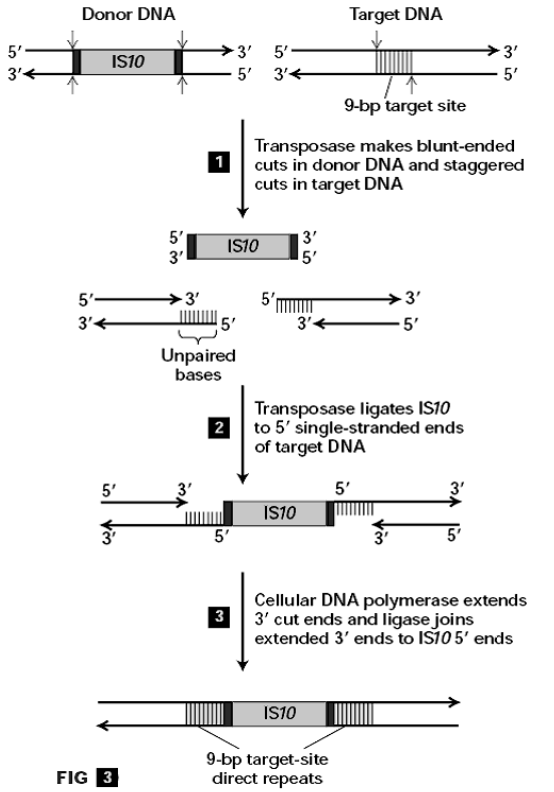


FIG 3
Model for transposition of bacterial insertion sequences.

Eukaryotic DNA Transposons McClintock's original discovery of mobile elements came from observation of certain spontaneous mutations in maize that affect production of any of the several enzymes required to make anthocyanin, a purple pigment in maize kernels. Mutant kernels are white, and wild-type kernels are purple. One class of these mutations is revertible at high frequency, whereas a second class of mutations does not revert unless they occur in the presence of the first class of mutations. McClintock called the agent responsible for the first class of mutations the *activator (Ac) element* and those responsible for the second class *dissociation (Ds) elements* because they also tended to be associated with chromosome breaks.

Many years after McClintock's pioneering discoveries, cloning and sequencing revealed that Ac elements are equivalent to bacterial IS elements. Like IS elements, they contain inverted terminal repeat sequences that flank the coding region for a transposase, which recognizes the terminal repeats and catalyzes transposition to a new site in DNA. Ds elements are deleted forms of the Ac element in which a portion of the sequence encoding transposase is missing. Because it does not encode a functional transposase, a Ds element cannot move by itself. However, in plants that carry the Ac element, and thus express a functional transposase, Ds elements can move. Since McClintock's early work on mobile elements in corn, transposons have been identified in other eukaryotes.

For instance, approximately half of all the spontaneous mutations observed in *Drosophila* are due to the insertion of mobile elements. Although most of the mobile elements in *Drosophila* function as retrotransposons, at least one—the **P element**—functions as a DNA transposon, moving by a cut-and-paste mechanism similar to that used by bacterial insertion sequences. Current methods for constructing transgenic *Drosophila* depend on engineered, high-level expression of the P-element transposase and use of the P-element inverted terminal repeats as targets for transposition.

DNA transposition by the cut-and-paste mechanism can result in an increase in the copy number of a transposon when it occurs during S phase, the period of the cell cycle when DNA synthesis occurs. This happens when the donor DNA is from one of the two daughter DNA molecules in a region of a chromosome that has replicated and the target DNA is in the region that has not yet replicated. When DNA replication is complete at the end of the S phase, the target DNA in its new location is also replicated. This results in a net increase by one in the total number of these transposons in the cell. When this occurs during the S phase preceding meiosis, two of the four germ cells produced have the extra copy. Repetition of this process over evolutionary time has resulted in the accumulation of large numbers of DNA transposons in the genomes of some organisms. Human DNA contains about 300,000 copies of full-length and deleted DNA tranposons, amounting to ≈3 percent of human DNA.

Some Retrotransposons Contain LTRs and Behave Like Intracellular Retroviruses

The genomes of all eukaryotes studied from yeast to humans contain retrotransposons, mobile DNA elements that transpose through an RNA intermediate utilizing a reverse transcriptase (see Figure 1b). These mobile elements are divided into two major categories, those containing and those lacking **long terminal repeats (LTRs).** LTR retrotransposons, which we discuss in this section, are common in yeast (e.g., Ty elements) and in *Drosophila* (e.g., *copia* elements). Although less abundant in mammals than non-LTR retrotransposons, LTR retrotransposons nonetheless constitute ≈8 percent of human genomic DNA. Because they exhibit some similarities with retroviruses,

these mobile elements sometimes are called *viral retrotransposons*. In mammals, retrotransposons lacking LTRs are the most common type of mobile element.

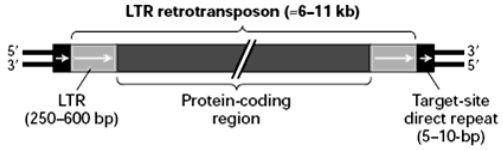


FIG 4 General structure of eukaryotic LTR retrotransposons.

The general structure of LTR retrotransposons found in eukaryotes is depicted in Figure 4. In addition to short 5′ and 3′ direct repeats typical of all mobile elements, these retrotransposons are marked by the presence of LTRs flanking the central protein-coding region. These *l*ong direct *t*erminal *r*epeats, containing ≈250–600 base pairs, are characteristic of integrated retroviral DNA and are critical to the life cycle of retroviruses. In addition to sharing LTRs with retroviruses, LTR retrotransposons encode all the proteins of
the most common type of retroviruses, except for the envelope proteins. Lacking these envelope proteins, LTR retrotransposons cannot bud from their host cell and infect other cells; however, they can transpose to new sites in the DNA of their host cell.

A key step in the retroviral life cycle is formation of retroviral genomic RNA from integrated retroviral DNA. This process serves as a model for generation of the RNA intermediate during transposition of LTR retrotransposons. As depicted in Figure 5, the leftward retroviral LTR functions as a promoter that directs host-cell RNA polymerase II to initiate transcription at the 5′ nucleotide of the R sequence. After the entire downstream retroviral DNA has been transcribed, the RNA sequence corresponding to the rightward LTR directs host-cell RNA-processing enzymes to cleave the primary transcript and add a poly(A) tail at the 3′end of the R sequence.

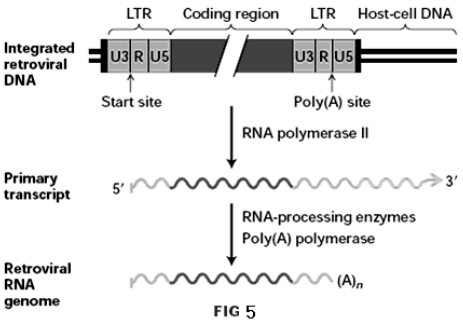


FIG 5
Generation of retroviral genomic RNA from integrated retroviral DNA.

The resulting retroviral RNA genome, which lacks a complete LTR, is packaged into a virion that buds from the host cell. After a retrovirus infects a cell, reverse transcription of its RNA genome by the retrovirus-encoded reverse transcriptase yields a double-stranded DNA containing complete LTRs (Figure 6).

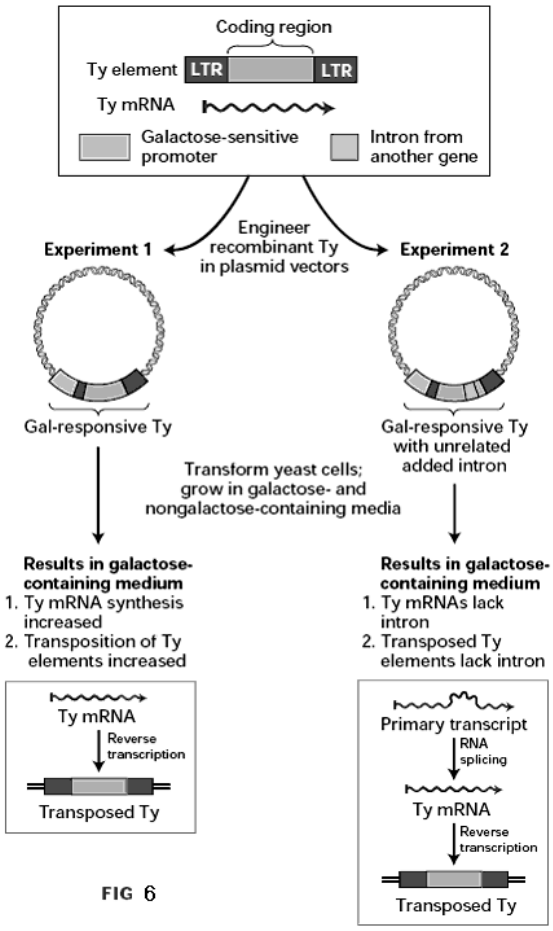
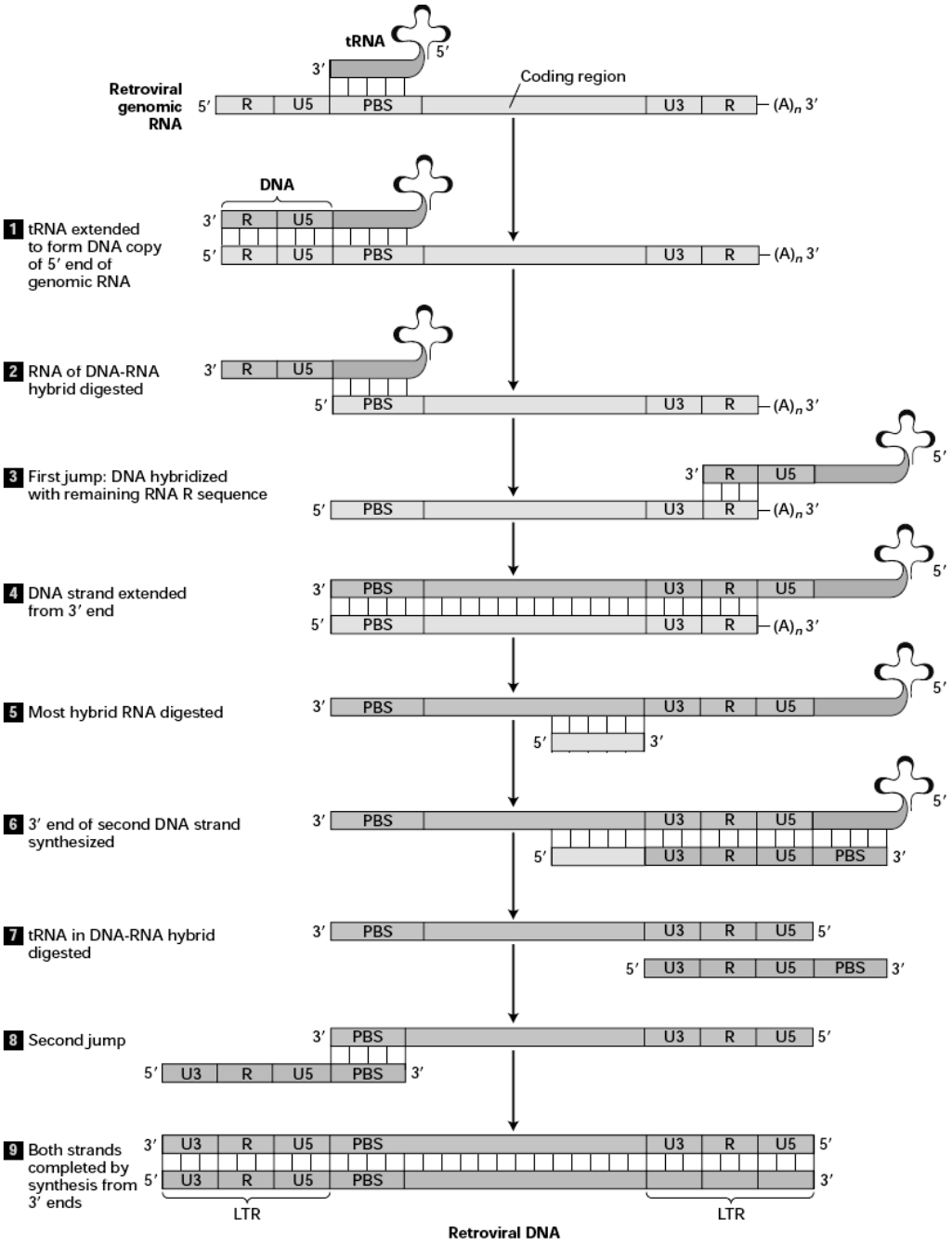


FIG 6

Integrase, another enzyme encoded by retroviruses that is closely related to the transposase of some DNA transposons, uses a similar mechanism to insert the double-stranded retroviral DNA into the host cell genome. In this process, short direct repeats of the target site sequence are generated at either end of the inserted viral DNA sequence.

As noted above, LTR retrotransposons encode reverse transcriptase and integrase. By analogy with retroviruses, these mobile elements are thought to move by a "copy and paste" mechanism whereby reverse transcriptase converts an RNA copy of a donor-site element into DNA, which is inserted into a target site by integrase. The experiments depicted in Figure 6 provided strong evidence for the role of an RNA intermediate in transposition of Ty elements.

Sequencing of the human genome has revealed that the most common LTR retrotransposon–related sequences in humans are derived from endogenous retroviruses (ERVs). Most of the 443,000 ERV-related DNA sequences in the human genome consist only of isolated LTRs. These are derived from full-length proviral DNA by homologous recombination between the two LTRs, resulting in deletion of the internal retroviral sequences.



Retrotransposons That Lack LTRs Move by a Distinct Mechanism

The most abundant mobile elements in mammals are retrotransposons that lack LTRs, sometimes called *nonviral retrotransposons*. These moderately repeated DNA sequences form two classes in mammalian genomes: *long interspersed elements (LINEs)* and *short interspersed elements (SINEs)*. In humans, full-length LINEs are ≈6 kb long, and SINEs are ≈300 bp long. Repeated sequences with the characteristics of LINEs have been observed in protozoans, insects, and plants, but for unknown reasons they are particularly abundant in the genomes of mammals. SINEs also are found primarily in mammalian DNA. Large numbers of LINEs and SINEs in higher eukaryotes have accumulated over evolutionary time by repeated copying of sequences at a few positions in the genome and insertion of the copies into new positions. Although these mobile elements do not contain LTRs, the available evidence indicates that they transpose through an RNA intermediate.

LINEs: Human DNA contains three major families of LINE sequences that are similar in their mechanism of transposition, but differ in their sequences: L1, L2, and L3. Only members of the L1 family transpose in the contemporary human genome. LINE sequences are present at ≈900,000 sites in the human genome, accounting for a staggering 21 percent of total human DNA. The general structure of a complete LINE is diagrammed in Figure 8. LINEs usually are flanked by short direct repeats, the hallmark of mobile elements, and contain two long open reading frames (ORFs). ORF1, ≈1 kb long, encodes an RNA-

binding protein. ORF2, ≈4 kb long, encodes a protein that has a long region of homology with the reverse transcriptases of retroviruses and viral retrotransposons, but also exhibits DNA endonuclease activity.

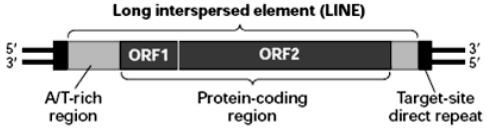


FIG 8 General structure of a LINE, one of the two classes of non-LTR retrotransposons in mammalian DNA.

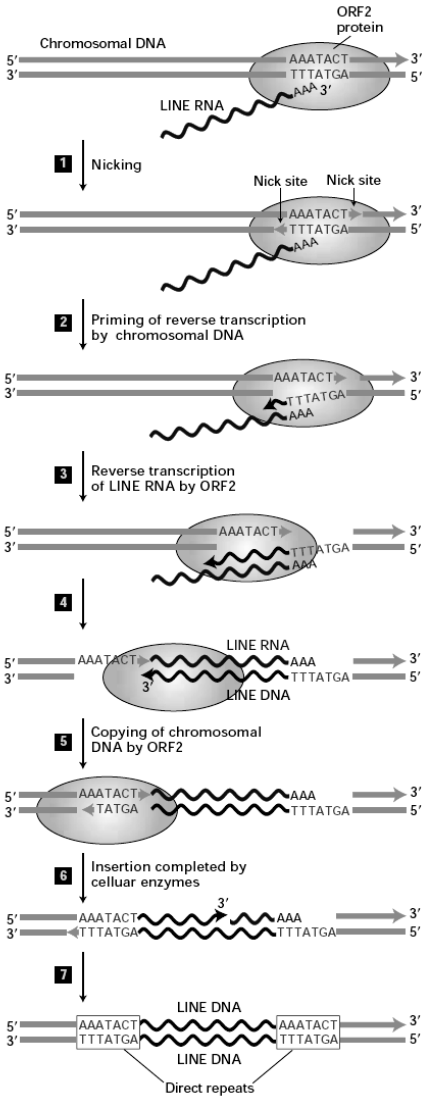
Evidence for the mobility of L1 elements first came from analysis of DNA cloned from humans with certain genetic diseases. DNA from these patients was found to carry mutations resulting from insertion of an L1 element into a gene, whereas no such element occurred within this gene in either parent. About 1 in 600 mutations that cause significant disease in humans are due to L1 transpositions or SINE transpositions that are catalyzed by L1- encoded proteins. Later experiments similar to those just described with yeast Ty elements (see Figure 6) confirmed that L1 elements transpose through an RNA intermediate. In these experiments, an intron was introduced into a cloned mouse L1 element, and the recombinant L1 element was stably transformed into cultured hamster cells. After several cell doublings, a PCR-amplified fragment corresponding to the L1 element but lacking the inserted intron was detected in the cells. This finding strongly suggests that over time the recombinant L1 element containing the inserted intron had transposed to new sites in the hamster genome through an RNA intermediate that underwent RNA splicing to remove the intron.

Since LINEs do not contain LTRs, their mechanism of transposition through an RNA intermediate differs from that of LTR retrotransposons. ORF1 and ORF2 proteins are translated from a LINE RNA. In vitro studies indicate that transcription by RNA polymerase II is directed by promoter sequences at the left end of integrated LINE DNA. LINE RNA is polyadenylated by the same post-transcriptional mechanism that polyadenylates other mRNAs. The LINE RNA then is transported into the cytoplasm, where it is translated into ORF1 and ORF2 proteins. Multiple copies of ORF1 protein then bind to the LINE RNA, and ORF2 protein binds to the poly(A) tail.

The LINE RNA is then transported back into the nucleus as a complex with ORF1 and ORF2. ORF2 then makes staggered nicks in chromosomal DNA on either side of any A/Trich sequence in the genome (Figure 9, step 1). Reverse transcription of LINE RNA by ORF2 is primed by the single-stranded T-rich sequence generated by the nick in the bottom strand, which hybridizes to the LINE poly(A) tail (step 2). ORF2 then reverse-transcribes the LINE RNA (step 3) and then continues this new DNA strand, switching to the single-stranded region of the upper chromosomal strand as a template (steps 4 and 5). Cellular enzymes then hydrolyze the RNA and extend the 3_ end of the chromosomal DNA top strand, replacing the LINE RNA strand with DNA (step 6).

Finally, 5’ and 3’ ends of DNA strands are ligated, completing the insertion (step 7). These last steps (6 and 7) probably are catalyzed by the same cellular enzymes that remove RNA primers and ligate Okazaki fragments during DNA replication. The complete process results in insertion of a copy of the original LINE retrotransposon into a new site in chromosomal DNA. A short direct repeat is generated at the insertion site because of the initial staggered cleavage of the two chromosomal DNA strands (step 1).

The vast majority of LINEs in the human genome are truncated at their 5’ end, suggesting that reverse transcription terminated before completion and the resulting fragments extending variable distances from the poly(A) tail were inserted. Because of this shortening, the average size of LINE elements is only about 900 base pairs, whereas the full-length sequence is ≈6 kb long. In addition, nearly all the full-length elements contain stop codons and frameshift mutations in ORF1 and ORF2; these mutations probably have accumulated in most LINE sequences over evolutionary time. As a result of truncation and mutation, only ≈0.01 percent of the LINE sequences in the human genome are full-length with intact open reading frames for ORF1 and ORF2, ≈60–100 in total.



SINEs: The second most abundant class of mobile elements in the human genome, SINEs constitute ≈13 percent of total human DNA. Varying in length from about 100 to 400 base pairs, these retrotransposons do not encode protein, but most contain a 3’ A/T-rich sequence similar to that in LINEs. SINEs are transcribed by RNA polymerase III, the same nuclear RNA polymerase that transcribes genes encoding tRNAs, 5S rRNAs, and other small stable RNAs . Most likely, the ORF1 and ORF2 proteins expressed from full-length LINEs mediate transposition of SINEs by the retrotransposition mechanism depicted in Figure 9.

SINEs occur at about 1.6 million sites in the human genome. Of these, ≈1.1 million are *Alu elements,* so named because

most of them contain a single recognition site for the restriction enzyme *Alu*I. *Alu* elements exhibit considerable sequence homology with and may have evolved from 7SL RNA, a component of the signal-recognition particle. This abundant cytosolic ribonucleoprotein particle aids in targeting certain polypeptides, as they are being synthesized, to the membranes of the endoplasmic reticulum. *Alu* elements are scattered throughout the human genome at sites where their insertion has not disrupted gene expression: between genes, within introns, and in the 3' untranslated regions of some mRNAs. For instance, nine *Alu* elements are located within the human β-globin gene cluster. The overall frequency of L1 and SINE retrotranspositions in humans is estimated to be about one new retrotransposition in very eight individuals, with ≈40 percent being L1 and 60 percent SINEs, of which ≈90 percent are *Alu* elements.

Similar to other mobile elements, most SINEs have accumulated mutations from the time of their insertion in the germ line of an ancient ancestor of modern humans. Like LINEs, many SINEs also are truncated at their 5' end. Table summarizes the major types of interspersed repeats derived from mobile elements in the human genome.

Class	Length	Copy Number in Human Genome	Fraction of Human Genome %
DNA transposons	2–3 kb	300,000	3
LTR retrotransposons	6–11 kb	440,000	8
Non-LTR retrotransposons			
LINEs	6–8 kb	860,000	21
SINEs	100–300 bp	1,600,000	13

In addition to the mobile elements listed in Table, DNA copies of a wide variety of mRNAs appear to have integrated into chromosomal DNA. Since these sequences lack introns and do not have flanking sequences similar to those of the functional gene copies, they clearly are not simply duplicated genes that have drifted into nonfunctionality and become pseudogenes, as discussed earlier.

Instead, these DNA segments appear to be retrotransposed copies of spliced and polyadenylated (processed) mRNA. Compared with normal genes encoding mRNAs, these inserted segments generally contain multiple mutations, which are thought to have accumulated since their mRNAs were first reverse-transcribed and randomly integrated into the genome of a germ cell in an ancient ancestor. These nonfunctional genomic copies of mRNAs are referred to as *processed pseudogenes.* Most processed pseudogenes are flanked by short direct repeats, supporting the hypothesis that they were generated by rare retrotransposition events involving cellular mRNAs.

Other moderately repetitive sequences representing partial or mutant copies of genes encoding small nuclear RNAs (snRNAs) and tRNAs are found in mammalian genomes. Like processed pseudogenes derived from mRNAs, these nonfunctional copies of small RNA genes are flanked by short direct repeats and most likely result from rare retrotransposition events that have accumulated through the course of evolution. Enzymes expressed from a LINE are thought to have carried out all these retrotransposition events involving mRNAs, snRNAs, and tRNAs.

Mobile DNA Elements Probably Had a Significant Influence on Evolution

Although mobile DNA elements appear to have no direct function other than to maintain their own existence, their presence probably had a profound impact on the evolution of modern-day organisms. As mentioned earlier, about half the spontaneous mutations in *Drosophila* result from insertion of a mobile DNA element into or near a transcription unit. In mammals, however, mobile elements cause a much smaller proportion of spontaneous mutations: ≈10 percent in mice and only 0.1–0.2 percent in humans. Still, mobile elements have been found in mutant alleles associated with several human genetic diseases.

In lineages leading to higher eukaryotes, homologous recombination between mobile DNA elements dispersed throughout ancestral genomes may have generated gene duplications and other DNA rearrangements during evolution. For instance, cloning and sequencing of the β-globin gene cluster from various primate species has provided strong evidence that the human G_γ and A_γ genes arose from an unequal homologous crossover between two L1 sequences flanking an ancestral globin gene. Subsequent divergence of such duplicated genes could lead to acquisition of distinct, beneficial functions associated with each member of a gene family. Unequal crossing over between mobile elements located within introns of a particular gene could lead to the duplication of exons within that gene. This process most likely influenced the evolution of genes that contain multiple copies of similar exons encoding similar protein domains, such as the fibronectin gene.

Some evidence suggests that during the evolution of higher eukaryotes, recombination between interspersed repeats in introns of *two separate* genes also occurred, generating new genes made from novel combinations of preexisting exons (Figure 10).

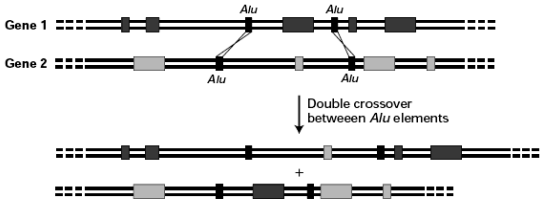


FIG 10 Exon shuffling via recombination between homologous interspersed repeats.

This evolutionary process, termed **exon shuffling,** may have occurred during evolution of the genes encoding tissue plasminogen activator, the Neu receptor, and epidermal growth factor, which all contain an EGF domain. In this case, exon shuffling presumably resulted in insertion of an EGF domain–encoding exon into an intron of the ancestral form of each of these genes. Both DNA transposons and LINE retrotransposons have been shown to occasionally carry unrelated flanking sequences when they insert into new sites by the mechanisms diagrammed in Figure 11.

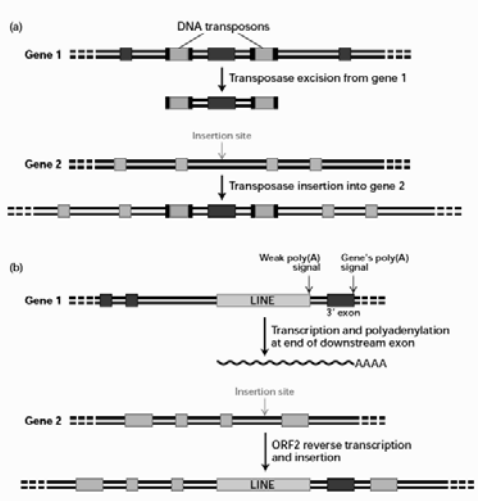


FIGURE 11 Exon shuffling by transposition.

These mechanisms likely also contributed to exon shuffling during the evolution of contemporary genes.

In addition to causing changes in coding sequences in the genome, recombination between mobile elements and transposition of DNA adjacent to DNA transposons and retrotransposons likely played a significant role in the evolution of regulatory sequences that control gene expression. As noted earlier, eukaryotic genes have transcription-control regions, called enhancers, that can operate over distances of tens of thousands of base pairs. Transcription of many genes is controlled through the combined effects of several enhancer elements. Insertion of mobile elements near such transcription-control regions probably contributed to the evolution of new combinations of enhancer sequences. These in turn control which specific genes are expressed in particular cell types and the amount of the encoded protein produced in modern organisms.

These considerations suggest that the early view of mobile DNA elements as completely selfish molecular parasites misses the mark. Rather, they have likely contributed profoundly to the evolution of higher organisms by promoting (1) the generation of gene families via gene duplication, (2) the creation of new genes via shuffling of preexisting exons, and (3) formation of more complex regulatory regions that provide multifaceted control of gene expression.

D. Cell Division and Cell Cycle

I. Systems of Cell Division

1. Four events occur before and during cell division.
- A signal to reproduce must be received.
- Replication of DNA and vital cell components must occur.
- DNA must distribute to the new cells.
- The cell membrane (and cell wall in some organisms) must separate the two new cells.

II. Prokaryotes divide by fission

1. Prokaryotic cells grow in size, replicate DNA, and divide into two new cells. This process is called *fission*. *Escherichia coli* (a bacterium) simply divides as quickly as resources permit. At 37°C, this is about once every 40 minutes. When resources are abundant, *E. coli* can divide every 20 minutes.
    - Prokaryotes generally have just one circular chromosome.
    - The *E. coli* chromosome is 1.6 mm in diameter, making the unfolded circle 100 times greater than the size of the cell. The molecule is packaged by folding in on itself with the aid of basic proteins that associate with the acidic DNA.
    - Circular chromosomes appear to be characteristic of all prokaryotes.
3. The prokaryotes have a site called *ori*, where DNA replication begins, and a site *ter*, where it ends.
    - *Ori* is short for origin of replication.
    - *Ter* is short for terminus of replication.
4. As DNA replicates, each of the two resulting DNA molecules attaches to the plasma membrane. As the bacterium grows, new plasma membrane is added between the attachment points, and the DNA molecules are moved apart.
5. Cytokinesis, which is cell partitioning, begins around 20 minutes after chromosome duplication is completed. A pinching of the plasma membrane to form a constricting ring separates the one cell into two, each with a complete chromosome.
    - A tubulin-like fiber is involved in the purse-string constriction.

III. Eukaryotic cells divide by mitosis or meiosis

1. All reproduction involves reproduction signals, DNA replication, segregation, and cytokinesis.
2. Unlike prokaryotes, eukaryotic cells do not constantly divide whenever environmental conditions are adequate, although unicellular eukaryotes do so more often than the cells of multicellular organisms.
    - Some differentiated cells of multicellular organisms rarely or never divide.
    - Signals to divide are related to the needs of the entire organism, not simply the opportunity created by resources.
3. Eukaryotes usually have many chromosomes. Eukaryotes have a nucleus, which must replicate and, with few exceptions, divide during cell division. Mitosis generates two cells with the same genetic information as the original cell.
4. Meiosis is a specialized cell division used for sexual reproduction. The genetic information of the chromosomes is shuffled, and the cells, called gametes, typically get one-half of the original DNA complement.

D1. Mitosis

Mitosis is the process by which a eukaryotic cell separates the chromosomes in its cell nucleus, into two identical sets in two daughter nuclei. It is generally followed immediately by cytokinesis, which divides the nuclei, cytoplasm, organelles and cell membrane into two daughter cells containing roughly equal shares of these cellular components. Mitosis and cytokinesis together define the mitotic (M) phase of the cell cycle - the division of the mother cell into two daughter cells, genetically identical to each other and to their parent cell.

Mitosis occurs exclusively in eukaryotic cells, but occurs in different ways in different species. For example, animals undergo an "open" mitosis, where the nuclear envelope breaks down before the chromosomes separate, while fungi such as Aspergillus nidulans and Saccharomyces cerevisiae (yeast) undergo a "closed" mitosis, where chromosomes divide within an intact cell nucleus.[2] Prokaryotic cells, which lack a nucleus, divide by a process called binary fission.

The process of mitosis is complex and highly regulated. The sequence of events is divided into phases, corresponding to the completion of one set of activities and the start of the next. These stages are prophase, prometaphase, metaphase, anaphase and telophase. During the process of mitosis the pairs of chromosomes condense and attach to fibers that pull the sister chromatids to opposite sides of the cell. The cell then divides in cytokinesis, to produce two identical daughter cells.

Because cytokinesis usually occurs in conjunction with mitosis, "mitosis" is often used interchangeably with "mitotic phase". However, there are many cells where

mitosis and cytokinesis occur separately, forming single cells with multiple nuclei. This occurs most notably among the fungi and slime moulds, but is found in various different groups. Even in animals, cytokinesis and mitosis may occur independently, for instance during certain stages of fruit fly embryonic development.[4] Errors in mitosis can either kill a cell through apoptosis or cause mutations that may lead to cancer.

Overview

The primary result of mitosis is the division of the parent cell's genome into two daughter cells. The genome is composed of a number of chromosomes, complexes of tightly-coiled DNA that contain genetic information vital for proper cell function. Because each resultant daughter cell should be genetically identical to the parent cell, the parent cell must make a copy of each chromosome before mitosis. This occurs during S phase, in interphase, the period that precedes the mitotic phase in the cell cycle where preparation for mitosis occurs.

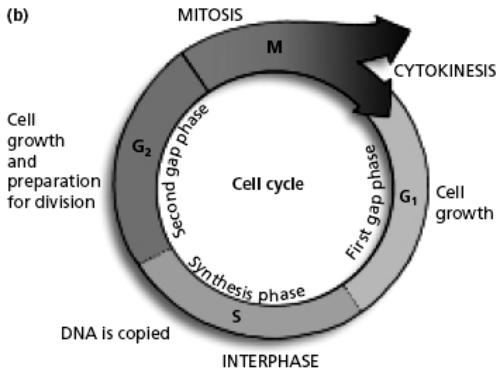
Each new chromosome now contains two identical copies of itself, called sister chromatids, attached together in a specialized region of the chromosome known as the centromere. Each sister chromatid is not considered a chromosome in itself, and a chromosome does not always contain two sister chromatids.

In most eukaryotes, the nuclear envelope that separates the DNA from the cytoplasm disassembles. The chromosomes align themselves in a line spanning the cell. Microtubules, essentially miniature strings, splay out from opposite ends of the cell and shorten, pulling apart the sister chromatids of each chromosome. As a matter of convention, each sister chromatid is now considered a chromosome, so they are renamed to sister chromosomes. As the cell elongates, corresponding sister chromosomes are pulled toward opposite ends. A new nuclear envelope forms around the separated sister chromosomes.

As mitosis completes cytokinesis is well underway. In animal cells, the cell pinches inward where the imaginary line used to be, (the pinching of the cell membrane to form the two daughter cells is called cleavage furrow) separating the two developing nuclei. In plant cells, the daughter cells will construct a new dividing cell wall between each other. Eventually, the mother cell will be split in half, giving rise to two daughter cells, each with an equivalent and complete copy of the original genome.

Prokaryotic cells undergo a process similar to mitosis called binary fission. However, prokaryotes cannot be properly said to undergo mitosis because they lack a nucleus and only have a single chromosome with no centromere.

Phases of cell cycle and mitosis



Interphase

The mitotic phase is a relatively short period of the cell cycle. It alternates with the much longer interphase, where the cell prepares itself for cell division. Interphase is divided into three phases, G1 (first gap), S (synthesis), and G2 (second gap). During all three phases, the cell grows by producing proteins and cytoplasmic organelles. However, chromosomes are replicated only during the S phase. Thus, a cell grows (G1), continues to grow as it duplicates its chromosomes (S), grows more and prepares for mitosis (G2), and divides (M).

Preprophase

In plant cells only, prophase is preceded by a pre-prophase stage. In highly vacuolated plant cells, the nucleus has to migrate into the center of the cell before mitosis can begin. This is achieved through the formation of a phragmosome, a transverse sheet of cytoplasm that bisects the cell along the future plane of cell division. In addition to phragmosome formation, preprophase is characterized by the formation of a ring of microtubules and actin filaments (called preprophase band) underneath the plasmamembrane around the equatorial plane of the future mitotic spindle and predicting the position of cell plate fusion during telophase. The cells of higher plants (such as the flowering plants) lack centrioles. Instead, spindle microtubules aggregate on the surface of the nuclear envelope during prophase. The preprophase band disappears during nuclear envelope disassembly and spindle formation in prometaphase.

Prophase

Normally, the genetic material in the nucleus is in a loosely bundled coil called chromatin. At the onset of prophase, chromatin condenses together into a highly ordered structure called a chromosome. Since the genetic material has already been duplicated earlier in S phase, the replicated chromosomes have two sister chromatids, bound together at the centromere by the cohesion complex. Chromosomes are visible at high magnification through a light microscope.

Close to the nucleus are two centrosomes. Each centrosome, which was replicated earlier independent of mitosis, acts as a coordinating center for the cell's microtubules. The two centrosomes nucleate microtubules (which may be thought of as cellular ropes or poles) by polymerizing soluble tubulin present in the cytoplasm. Molecular motor proteins create repulsive forces that will push the centrosomes to opposite side of the nucleus. The centrosomes are only present in animals. In plants the microtubules form independently.

Some centrosomes contain a pair of centrioles that may help organize microtubule assembly, but they are not essential to formation of the mitotic spindle.

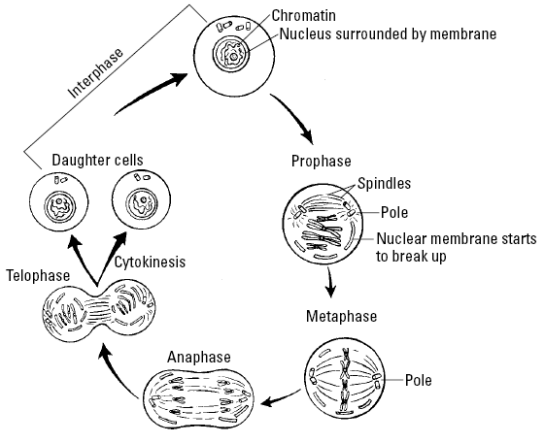
Prometaphase

The nuclear envelope disassembles and microtubules invade the nuclear space. This is called open mitosis, and it occurs in most multicellular organisms. Fungi and some protists, such as algae or trichomonads, undergo a variation called closed mitosis where the spindle forms inside the nucleus or its microtubules are able to penetrate an intact nuclear envelope.

Each chromosome forms two kinetochores at the centromere, one attached at each chromatid. A kinetochore is a complex protein structure that is analogous to a ring for the microtubule hook; it is the point where microtubules attach themselves to the chromosome. Although the kinetochore structure and function are not fully understood, it is known that it contains some form of molecular motor.

When a microtubule connects with the kinetochore, the motor activates, using energy from ATP to "crawl" up the tube toward the originating centrosome. This motor activity, coupled with polymerisation and depolymerisation of microtubules, provides the pulling force necessary to later separate the chromosome's two chromatids.

When the spindle grows to sufficient length, kinetochore microtubules begin searching for kinetochores to attach to. A number of nonkinetochore microtubules find and interact with corresponding nonkinetochore microtubules from the opposite centrosome to form the mitotic spindle. Prometaphase is sometimes considered part of prophase.



Metaphase

As microtubules find and attach to kinetochores in prometaphase, the centromeres of the chromosomes convene along the metaphase plate or equatorial plane, an imaginary line that is equidistant from the two centrosome poles. This even alignment is due to the counterbalance of the pulling powers generated by the opposing kinetochores, analogous to a tug-of-war between people of equal strength. In certain types of cells, chromosomes do not line up at the metaphase plate and instead move back and forth between the poles randomly, only roughly lining up along the midline. Metaphase comes from the Greek μετα meaning "after."

Because proper chromosome separation requires that every kinetochore be attached to a bundle of microtubules (spindle fibres) , it is thought that unattached kinetochores generate a signal to prevent premature progression to anaphase without all chromosomes being aligned. The signal creates the mitotic spindle checkpoint.

Anaphase

When every kinetochore is attached to a cluster of microtubules and the chromosomes have lined up along the metaphase plate, the cell proceeds to anaphase (from the Greek ανα meaning "up," "against," "back," or "re-").

Two events then occur; First, the proteins that bind sister chromatids together are cleaved, allowing them to separate. These sister chromatids turned sister chromosomes are pulled apart by shortening kinetochore microtubules and move toward the respective centrosomes to which they are attached. Next, the nonkinetochore microtubules elongate, pushing the centrosomes (and the set of chromosomes to which they are attached) apart to opposite ends of the cell. The force that causes the centrosomes to move towards the ends of the cell is still unknown, although there is a theory that suggests that the rapid assembly and breakdown of microtubules may cause this movement.

These two stages are sometimes called early and late anaphase. Early anaphase is usually defined as the separation of the sister chromatids, while late anaphase is the elongation of the microtubules and the microtubules being pulled farther apart. At the end of anaphase, the cell has succeeded in separating identical copies of the genetic material into two distinct populations

Telophase

Telophase (from the Greek τελος meaning "end") is a reversal of prophase and prometaphase events. It "cleans up" the after effects of mitosis. At telophase, the nonkinetochore microtubules continue to lengthen, elongating the cell even more. Corresponding sister chromosomes attach at opposite ends of the cell. A new nuclear envelope, using fragments of the parent cell's nuclear membrane, forms around each set of separated sister chromosomes. Both sets of chromosomes, now surrounded by new nuclei, unfold back into chromatin. Mitosis is complete, but cell division is not yet complete.

Cytokinesis

Cytokinesis is often mistakenly thought to be the final part of telophase, however cytokinesis is a separate process that begins at the same time as telophase. Cytokinesis is technically not even a phase of mitosis, but rather a separate process, necessary for completing cell division. In animal cells, a cleavage furrow (pinch) containing a contractile ring develops where the metaphase plate used to be, pinching off the separated nuclei. In both animal and plant cells, cell division is also driven by vesicles derived from the Golgi apparatus, which move along microtubules to the middle of the cell. In plants this structure coalesces into a cell plate at the center of the phragmoplast and develops into a cell wall, separating the two nuclei. The phragmoplast is a microtubule structure typical for higher plants, whereas some green algae use a phycoplast microtubule array during cytokinesis. Each daughter cell has a complete copy of the genome of its parent cell. The end of cytokinesis marks the end of the M-phase.

Significance

The importance of mitosis is the maintenance of the chromosomal set; each cell formed receives chromosomes that are alike in composition and equal in number to the chromosomes of the parent cell. Transcription is generally believed to cease during mitosis, but epigenetic mechanisms such as bookmarking function during this stage of the cell cycle to ensure that the "memory" of which genes were active prior to entry into mitosis are transmitted to the daughter cells.

Consequences of errors

Although errors in mitosis are rare, the process may go wrong, especially during early cellular divisions in the zygote. Mitotic errors can be especially dangerous to the organism because future offspring from this parent cell will carry the same disorder.

In non-disjunction, a chromosome may fail to separate during anaphase. One daughter cell will receive both sister chromosomes and the other will receive none. This results in the former cell having three chromosomes coding for the same thing (two sisters and a homologue), a condition known as trisomy, and the latter cell having only one chromosome (the homologous chromosome), a condition known as monosomy. These cells are considered aneuploidic cells and these abnormal cells can cause cancer.

Mitosis is a traumatic process. The cell goes through dramatic changes in ultrastructure, its organelles

disintegrate and reform in a matter of hours, and chromosomes are jostled constantly by probing microtubules. Occasionally, chromosomes may become damaged. An arm of the chromosome may be broken and the fragment lost, causing deletion. The fragment may incorrectly reattach to another, non-homologous chromosome, causing translocation. It may reattach to the original chromosome, but in reverse orientation, causing inversion. Or, it may be treated erroneously as a separate chromosome, causing chromosomal duplication. The effect of these genetic abnormalities depend on the specific nature of the error. It may range from no noticeable effect, cancer induction, or organism death.

Endomitosis

Endomitosis is a variant of mitosis without nuclear or cellular division, resulting in cells with many copies of the same chromosome occupying a single nucleus. This process may also be referred to as endoreduplication and the cells as endoploid. An example of a cell that goes through endomitosis is the megakaryocyte and Yeast.

D2. Meiosis

In biology or life science, meiosis (pronounced my-oh-sis) is a process of reduction division in which the number of chromosomes per cell is cut in half. In animals, meiosis always results in the formation of gametes. The word "meiosis" comes from the Greek verb meioun, meaning "to make small," since it results in a reduction in chromosome number in the gamete cell.

Meiosis is essential for sexual reproduction and therefore occurs in all eukaryotes (including single-celled organisms) that reproduce sexually. A few eukaryotes, notably the Bdelloid rotifers, have lost the ability to carry out meiosis and have acquired the ability to reproduce by parthenogenesis. Meiosis does not occur in archaea or bacteria, which reproduce via asexual processes such as mitosis or binary fission. Each cell has half the number of chromosomes as the parent cell.

During meiosis, the genome of a diploid germ cell, which is composed of long segments of DNA packaged into chromosomes, undergoes DNA replication followed by two rounds of division, resulting in four haploid cells. Each of these cells contain one complete set of chromosomes, or half of the genetic content of the original cell. If meiosis produces gametes, these cells must fuse during fertilization to create a new diploid cell, or zygote before any new growth can occur. Thus, the division mechanism of meiosis is a reciprocal process to the joining of two genomes that occurs at fertilization. Because the chromosomes of each parent undergo genetic recombination during meiosis, each gamete, and thus each zygote, will have a unique genetic blueprint encoded in its DNA. Together, meiosis and fertilization constitute sexuality in the eukaryotes, and generate genetically distinct individuals in populations.

In all plants, and in many protists, meiosis results in the formation of haploid cells that can divide vegetatively without undergoing fertilization. In these groups, gametes are produced by mitosis.

Meiosis uses many of the same biochemical mechanisms employed during mitosis to accomplish the redistribution of chromosomes. There are several features unique to meiosis, most importantly the pairing and genetic recombination between homologous chromosomes.

History: Meiosis was discovered and described for the first time in sea urchin eggs in 1876, by noted German biologist Oscar Hertwig (1849-1922). It was described again in 1883, at the level of chromosomes, by Belgian zoologist Edouard Van Beneden (1846-1910), in Ascaris worms' eggs. The significance of meiosis for reproduction and inheritance, however, was described only in 1890 by German biologist August Weismann (1834-1914), who noted that two cell divisions were necessary to transform one diploid cell into four haploid cells if the number of chromosomes had to be maintained. In 1911 the American geneticist Thomas Hunt Morgan (1866-1945) observed crossover in Drosophila melanogaster meiosis and provided the first true genetics.

Evolution: Meiosis is thought to have appeared 1.4 billion years ago. The only supergroup of eukaryotes which does not have meiosis in all organisms is excavata. The other five major supergroups, opisthokonts, amoebozoa, rhizaria, archaeplastida and chromalveolates all seem to have genes for meiosis universally present, even if not always functional. Some excavata species do have meiosis which is consistent with the hypothesis that excavata is an ancient, paraphyletic grade. An example of eukaryotic organism in which meiosis does not exist is euglenoid.

Occurrence of meiosis in eukaryotic life cycles

Meiosis occur in eukaryotic life cycles involving sexual reproduction, comprising of the constant cyclical process of meiosis and fertilization. This takes place alongside normal mitotic cell division. In multicellular organisms, there is an intermediary step between the diploid and haploid transition where the organism grows. The organism will then produce the germ cells that continue in the life cycle. The rest of the cells, called somatic cells, function within the organism and will die with it.

Cycling meiosis and fertilisation events produces a series of transitions back and forth between alternating haploid and diploid states. The organism phase of the life cycle can occur either during the diploid state (gametic life cycle), or during the haploid state (zygotic life cycle), or both (sporic life cycle, in which there two distinct organism phases, one during the haploid state and the other during the diploid state). In this sense, there are three types of life cycles that utilize sexual reproduction, differentiated by the location of the organisms phase(s). In the gametic life cycle, the species is diploid, grown from a diploid cell called the zygote. In the zygotic life cycle the species is haploid instead, spawned by the proliferation and differentiation of a single haploid cell called the gamete. Humans, for example, are diploid creatures. Human stem cells undergo meiosis to create haploid gametes, which are spermatozoa for males or ova for females. These gametes then fertilize in the Fallopian tubes of the female, producing a diploid zygote. The zygote undergoes progressive stages of mitosis and differentiation, turns into a blastocyst and then gets implanted in the uterus endometrium to create an embryo.

In the gametic life cycle, of which humans are a part, the living organism is diploid in nature. Here, we will generalize the example of human reproduction stated previously. The organism's diploid germ-line stem cells undergo meiosis to create haploid gametes, which fertilize to form the zygote. The diploid zygote undergoes repeated cellular division by mitosis to grow into the organism. Mitosis is a related process to meiosis that creates two cells that are genetically

identical to the parent cell. The general principle is that mitosis creates somatic cells and meiosis creates germ cells.

In the zygotic life cycle, the living organism is haploid. Two organisms of opposing gender contribute their haploid germ cells to form a diploid zygote. The zygote undergoes meiosis immediately, creating four haploid cells. These cells undergo mitosis to create the organism. Many fungi and many protozoa are members of the zygotic life cycle.

Finally, in the sporic life cycle, the living organism alternates between haploid and diploid states. Consequently, this cycle is also known as the alternation of generations. The diploid organism's germ-line cells undergo meiosis to produce gametes. The gametes proliferate by mitosis, growing into a haploid organism. The haploid organism's germ cells then combine with another haploid organism's cells, creating the zygote. The zygote undergoes repeated mitosis and differentiation to become the diploid organism again. The sporic life cycle can be considered a fusion of the gametic and zygotic life cycles.

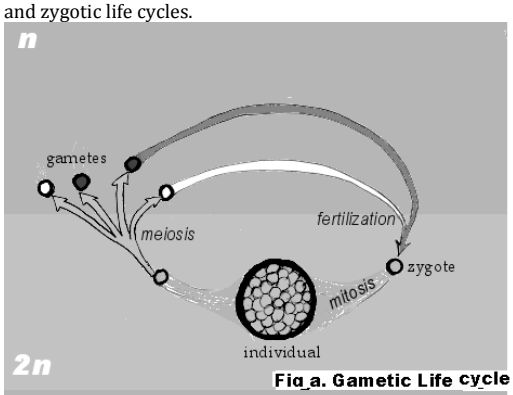


Fig a. Gametic Life cycle

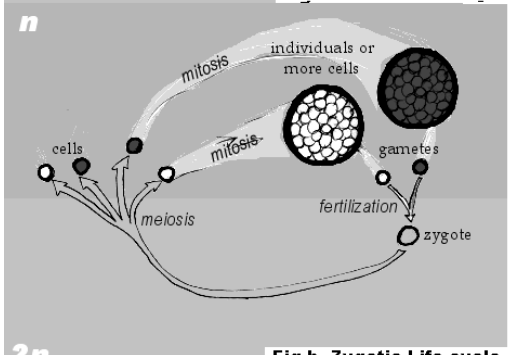


Fig b. Zygotic Life cycle

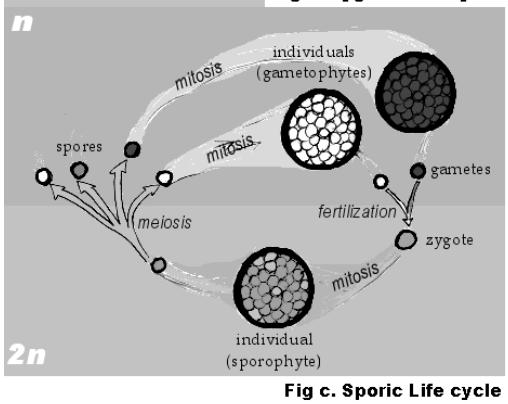


Fig c. Sporic Life cycle

Process: Because meiosis is a "one-way" process, it cannot be said to engage in a cell cycle as mitosis does. However, the preparatory steps that lead up to meiosis are identical in pattern and name to the interphase of the mitotic cell cycle.

Interphase is divided into three phases:

Growth 1 (G1) phase: Immediately follows cytokinesis. This is a very active period, where the cell synthesizes its vast array of proteins, including the enzymes and structural proteins it will need for growth. In G1 stage each of the 46 human chromosomes consists of a single (very long) molecule of DNA. At this point cells are 46,2N, identical to somatic cells.

Synthesis (S) phase: The genetic material is replicated: each of its chromosomes duplicates (46,2N). The cell is still diploid, however, because it still contains the same number of centromeres. However, the identical sister chromatids are in the chromatin form because spiralisation and condensation into denser chromosomes have not taken place yet. It will take place in prophase I in meiosis.

Growth 2 (G2) phase: G2 phase is absent in Meiosis

Interphase is immediately followed by meiosis I and meiosis II. Meiosis I consists of segregating the homologous chromosomes from each other, then dividing the diploid cell into two haploid cells each containing one of the segregates. Meiosis II consists of decoupling each chromosome's sister strands (chromatids), segregating the DNA into two sets of strands (each set containing one of each homologue), and dividing both haploid, duplicated cells to produce four haploid, unduplicated cells. Meiosis I and II are both divided into prophase, metaphase, anaphase, and telophase subphases, similar in purpose to their analogous subphases in the mitotic cell cycle. Therefore, meiosis encompasses the interphase (G1, S, G2), meiosis I (prophase I, metaphase I, anaphase I, telophase I), and meiosis II (prophase II, metaphase II, anaphase II, telophase II).

Meiosis generates genetic diversity in two ways: (1) independent assortment of chromosomes at both of the meiotic divisions allows genetic differences among gametes; and (2) physical exchange of chromosomal regions by homologous recombination during prophase I results in new genetic combinations within chromosomes.

Meiosis I

In meiosis I, the homologous pairs in a diploid cell separate , producing two haploid cells (46, N). The 46 chromosomes number is significant. A regular diploid cell contains 46 chromosomes and is considered 2N because it contains 23 pairs of homologous chromosomes. However, after meiosis I, although the cell contains 46 chromosomes it is only considered N because later in anaphase I the identical sister chromatids will remain together as the spindle pulls the pair toward the pole of the new cell. In meiosis II, a process similar to mitosis will occur whereby the sister chromatids are finally split, creating 2 haploid cells (23, N).

Prophase I

Homologous chromosomes pair and crossing over, or recombination, occurs--a step unique to meiosis. Chromosomes form structures called synapses. The paired chromosomes are called bivalents or tetrads, which have two chromosomes and four chromatids, with one chromosome coming from each parent. At this stage, non-sister chromatids may cross-over at points called chiasmata.

Leptotene: The first stage of prophase I is the leptotene stage, also known as leptonema, from Greek words meaning "thin threads." During this stage, individual chromosomes begin to condense into long strands within the nucleus. However the two sister chromatids are still so tightly bound that they are indistinguishable from one another. The chromosomes in the leptotene stage show a specific

arrangement where the telomeres are oriented towards the nuclear membrane. Hence, this stage is called "bouquet stage".

Zygotene: The zygotene stage, also known as zygonema, from Greek words meaning "paired threads,"occurs as the chromosomes approximately line up with each other into homologous chromosomes. The combined homologous chromosomes are said to be bivalent. They may also be referred to as a tetrad, a reference to the four sister chromatids. The two homologous chromosomes become "zipped" together, forming the synaptonemal complex, in a process known as synapsis.

Pachytene: The pachytene stage, also known as pachynema, from Greek words meaning "thick threads,"contains the following chromosomal crossover. Nonsister chromatids of homologous chromosomes randomly exchange segments of genetic information over regions of homology. (Sex chromosomes, however, are not identical, and only exchange information over a small region of homology.) Exchange takes place at sites where recombination nodules or chiasmata (singular: chiasma) have formed. The exchange of information between the non-sister chromatids results in a recombination of information; each chromosome has the complete set of information it had before, and there are no gaps formed as a result of the process. Because the chromosomes cannot be distinguished in the synaptonemal complex, the actual act of crossing over is not perceivable through the microscope.

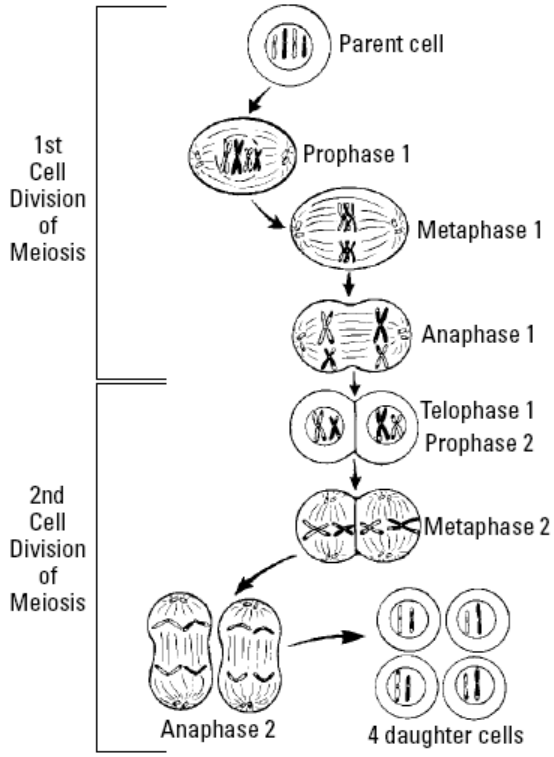
Diplotene: During the diplotene stage, also known as diplonema, from Greek words meaning "two threads," the synaptonemal complex degrades and homologous chromosomes separate from one another a little. The chromosomes themselves uncoil a bit, allowing some transcription of DNA. However, the homologous chromosomes of each bivalent remain tightly bound at chiasmata, the regions where crossing-over occurred. The chiasmata remain on the chromosomes until they are severed in Anaphase I.

In fetal oogenesis all developing oocytes develop to this stage and stop before birth. This suspended state is referred to as the dictyotene stage and remains so until puberty. In males, only spermatogonia exist until meiosis begins at puberty.

Diakinesis: Chromosomes condense further during the diakinesis stage, from Greek words meaning "moving through." This is the first point in meiosis where the four parts of the tetrads are actually visible. Sites of crossing over entangle together, effectively overlapping, making chiasmata clearly visible. Other than this observation, the rest of the stage closely resembles prometaphase of mitosis; the nucleoli disappear, the nuclear membrane disintegrates into vesicles, and the meiotic spindle begins to form.

Synchronous processes: During these stages, centrosomes, each containing a pair of centrioles are migrating to the two poles of the cell. These centrosomes, which were duplicated during S-phase, function as microtubule organizing centers nucleating microtubules, essentially cellular ropes and poles, during crossing over. They invade the nuclear membrane after it disintegrates, attaching to the chromosomes at the kinetochore. The kinetochore functions as a motor, pulling the chromosome along the attached microtubule toward the originating centriole, like a train on a track. There are four kinetochores on each tetrad, but the pair of kinetochores on each sister chromatid fuses and functions as a unit during meiosis I.

Microtubules that attach to the kinetochores are known as kinetochore microtubules. Other microtubules will interact with microtubules from the opposite centriole. These are also nonkinetochore microtubules.



Metaphase I

Homologous pairs move together along the phase plate: as kinetochore microtubules from both centrioles attach to their respective kinetochores, the homologous chromosomes align along an equatorial plane that bisects the spindle, due to continuous counterbalancing forces exerted on the bivalents by the microtubules emanating from the two kinetochores of homologous chromosomes. The physical basis of the independent assortment of chromosomes is the random orientation of each bivalent along the metaphase plate.

Anaphase I

Kinetochore microtubules shorten, severing the recombination nodules and pulling homologous chromosomes apart. Since each chromosome only has one functional unit of a pair of kinetochores[3], whole chromosomes are pulled toward opposing poles, forming two haploid sets. Each chromosome still contains a pair of sister chromatids. Nonkinetochore microtubules lengthen, pushing the centrioles further a part. The cell elongates in preparation for division down the middle.

Telophase I

The last meiotic division effectively ends when the centromeres arrive at the poles. Each daughter cell now has half the number of chromosomes but each chromosome consists of a pair of chromatids. This effect produces a variety of responses from the neuro-synchromatic enzyme, also known as NSE. The microtubules that make up the spindle network disappear, and a new nuclear membrane surrounds each haploid set. The chromosomes uncoil back into chromatin. Cytokinesis, the pinching of the cell

membrane in animal cells or the formation of the cell wall in plant cells, occurs, completing the creation of two daughter cells.

Cells enter a period of rest known as interkinesis or interphase II. No DNA replication occurs during this stage. Telophase I contains no nucleus, two daughter cells, and chromatids remain attached.

Meiosis II

Meiosis II is the second part of the meiotic process. Much of the process is similar to mitosis and meiosis I. End result is production of four haploid cells (23,1N) from the two haploid cells (46,1N) produced in meiosis I.

Prophase II takes an inversely proportional time compared to telophase I. In this prophase we see the disappearance of the nucleoli and the nuclear envelope again as well as the shortening and thickening of the chromatids. Centrioles move to the polar regions and arrange spindle fibers for the second meiotic division.

In metaphase II, the centromeres contain two kinetochores, that attach to spindle fibers from the centrosomes (centrioles) at each pole. The new equatorial metaphase plate is rotated by 90 degrees when compared to meiosis I, perpendicular to the previous plate.

This is followed by anaphase II, where the centromeres are cleaved, allowing microtubules attached to the kinetochores to pull the sister chromatids apart. The sister chromatids by convention are now called sister chromosomes as they move toward opposing poles.

The process ends with telophase II, which is similar to telophase I, and is marked by uncoiling and lengthening of the chromosomes and the disappearance of the microtubules. Nuclear envelopes reform and cleavage or cell wall formation eventually produces a total of four daughter cells, each with a haploid set of chromosomes. Meiosis is now complete.

The Significance of Meiosis

Meiosis facilitates stable sexual reproduction. Without the halving of ploidy, or chromosome count, fertilization would result in zygotes that have twice the number of chromosomes than the zygotes from the previous generation. Successive generations would have an exponential increase in chromosome count, resulting in an unwieldy genome that would cripple the reproductive fitness of the species. Polyploidy, the state of having three or more sets of chromosomes, also results in developmental abnormalities or lethality. Polyploidy is poorly tolerated in animal species. Plants, however, regularly produce fertile, viable polyploids. Polyploidy has been implicated as an important mechanism in plant speciation.

Most importantly, however, meiosis produces genetic variety in gametes that propagate to offspring. Recombination and independent assortment allow for a greater diversity of genotypes in the population. As a system of creating diversity, meiosis allows a species to maintain stability under environmental changes.

Nondisjunction

The normal separation of chromosomes in Meiosis I or sister chromatids in meiosis II is termed disjunction. When the separation is not normal, it is called nondisjunction. This results in the production of gametes which have either more or less of the usual amount of genetic material, and is a common mechanism for trisomy or monosomy. Nondisjunction can occur in the meiosis I or meiosis II, phases of cellular reproduction, or during mitosis.

This is a cause of several medical conditions in humans:

- Down Syndrome - trisomy of chromosome 21
- Patau Syndrome - trisomy of chromosome 13
- Edward Syndrome - trisomy of chromosome 18
- Klinefelter Syndrome - extra X chromosomes in males - ie XXY, XXXY, XXXXY
- Turner Syndrome - atypical X chromosome dosage in females - ie XO, XXX, XXXX
- XYY Syndrome - an extra Y chromosome in males

Meiosis in humans

In females, meiosis occurs in cells known as oogonia (singular: oogonium). Each oogonium that initiates meiosis will divide twice to form a single oocyte and three polar bodies. However, before these divisions occur, these cells stop at the diplotene stage of meiosis I and lay dormant within a protective shell of somatic cells called the follicle. Follicles begin growth at a steady pace in a process known as folliculogenesis, and a small number enter the menstrual cycle. Menstruated oocytes continue meiosis I and arrest at meiosis II until fertilization. The process of meiosis in females occurs during oogenesis, and differs from the typical meiosis in that it features a long period of meiotic arrest known as the Dictyate stage and lacks the assistance of centrosomes.

In males, meiosis occurs in precursor cells known as spermatogonia that divide twice to become sperm. These cells continuously divide without arrest in the seminiferous tubules of the testicles. Sperm is produced at a steady pace. The process of meiosis in males occurs during spermatogenesis.

D3 Cell Cycle

Fundamental concepts of Cell Cycle

- The **cell cycle** is the sequence of events between successive cell divisions.
- Many different processes must be coordinated during the cell cycle, some of which occur continuously(e.g. cell growth) and some discontinuously, as events or landmarks (e.g. cell division).
- Cell division must be coordinated with growth and DNA replication so that cell size and DNA content remain constant.
- The cell cycle comprises a **nuclear** or **chromosomal cycle** (DNA replication and partition) and a **cytoplasmic** or **cell division cycle** (doubling and division of cytoplasmic components, which in eukaryotes includes the organelles). The DNA is considered separately from other cell contents because it is usually present in only one or two copies per vegetative cell, and its replication and segregation must therefore be precisely controlled. Most of the remainder of the cell contents are synthesized continuously and in sufficient quantity to be distributed equally into the daughter cells when the parental cell is big enough to divide. An exception is the **centrosome,** an organelle that is pivotal in the process of chromosome segregation

itself, which is duplicated prior to mitosis and segregated into the daughter cells with the chromosomes (the **centrosome cycle).**

- In eukaryotes, the two major events of the chromosomal cycle, replication and mitosis, are controlled so that they can never occur simultaneously. Conversely, in bacteria the analogous processes, replication and partition, are coordinated so that partially replicated chromosomes can segregate during rapid growth. The eukaryotic cell cycle is divided into discrete phases which proceed in a particular order, whereas the stages **of** the bacterial cell cycle may overlap.
- The progress of the eukaryotic cell cycle is controlled at checkpoints where regulatory proteins receive input from monitors of the cell cycle itself (intrinsic information) and monitors of the environment (extrinsic information). Intrinsic monitoring insures that the stages of the cell cycle proceed in the correct order and that one stage **is** completed before the next begins.
- Extrinsic monitoring coordinates cell division with cell growth and arrests the cell cycle if the environment is unsuitable.
- The cell cycle is controlled by protein kinases. Cell cycle transitions involve positive feedback loops which cause sudden bursts of kinase activity, allowing switches in the states of phosphorylation of batteries of effector proteins. Cell cycle checkpoints are regulatory systems which inhibit those kinases if the internal or external environment is unsuitable. The alternation of DNA replication and mitosis is controlled by negative feedback - mitosis is inhibited by unfinished DNA replication, and DNA replication is prevented during mitosis by the phosphorylation and inactivation of a protein required for replication. The cell cycle is the result of a complex network of information, in which kinases are controlled by the integration of multiple positive and negative signals.

1 The bacterial cell cycle

DNA replication and growth coordination:

The **Helmstetter-Cooper model** (or I + C + D model) divides the bacterial chromosome cycle into three phases, the interval phase, the chromosome replication phase and the division phase, represented by the letters I, C and D, respectively. DNA replication occurs during the C phase; its duration is fixed (about 40 min in *E. coli),* reflecting the time taken to replicate the whole chromosome. The D phase begins when replication is complete, and culminates in cell division. The duration of the D phase is also fixed (about 20 min in E. *coli* and can be regarded as the time required to synthesize the cellular components required for cell division. The minimum duration of the chromosome cycle in *E.* coli is thus 1 h.

Because C + D is fixed, any change in the cell doubling time must reflect a change in the duration of I, the interval between successive initiations of replication. The doubling time of *E.* coli can be as long as 3 h or as short as 20 min. During slow growth, I > C + D and replication is completed before cell division. During rapid growth, however, the doubling time is shorter than the time taken to complete a round of replication and cell division. The only way the cell can accommodate its fixed chromosome cycle into the accelerated cell division cycle is to make I< C + D, i.e. new rounds of replication must begin before the previous round is complete. Therefore, during rapid growth, daughter cells inherit chromosomes which are already partially replicated (multiforked osomes) so that replication can be completed before the next round of cell division.

The frequency of initiation is thought to be controlled by a positive regulator which must be present at a certain critical concentration per *origin of* replication (q.v.) for initiation to be successful.

During rapid growth, the regulator accumulates more quickly, allowing more frequent initiation. Once initiation has occurred, the number of origins in the cell doubles and the effective concentration of the regulator is halved so that it must accumulate again for another round of initiation to occur. The existence of a positive regulator is predicted because *de novo* protein synthesis is required for initiation; however, the nature of this putative molecule is unknown. The replication initiator protein DnaA is a possible candidate, and factors which control methylation at *oriC* (and the *dna A* promoter) could also be involved (q.v. origin of replication, *Dam* methylation).

Partition and cytokinesis*:*

The partition of the replicated chromosome marks the culmination of the chromosome cycle and is followed by cell division. A septum forms at the midpoint of the parental cell, which is identified by a periseptal annulus, a region of modified cell envelope where the inner
and outer membranes are joined together around the circumference of the cell. Additional annuli form by duplication and migrate to positions equivalent to one-quarter and three-quarter cell lengths, and these are the sites of septation in daughter cells during the next round of cell division.

Once the septum has formed, the cell undergoes cytokinesis - it divides by binary fission. The identification of mutants which disrupt cell division or partitioning has shown that the two processes can be unhitched, and such mutants fall into several categories. *fts* mutants are deficient in septum formation and thus formfilaments that are temperature sensitive (hence the name). The filaments often contain regularly spaced nucleoids, indicating that replication and partitioning mechanisms are still functioning normally. *min* mutants generate septa too frequently, resulting in the formation of minicells, which are small cells which contain no chromosomal DNA (although they may contain plasmids). Finally, *pur* mutants form normal sized cells but fail to partition the chromosomes properly, so that diploid and anucleate cells arise with high frequency.

The pathway controlling cell division and partition has yet to be determined in full, but several key players have been identified. A good candidate for the initiator of cell division is FtsZ. This protein is structurally and functionally similar to tubulin, which forms the contractile ring in eukaryotic cells. It is distributed ubiquitously during most of the cell cycle, but is localized around the annulus at the beginning of the D phase as a Z ring. Its abundance appears to correlate exactly with the frequency of cell division, thusftsz mutants fail to form septa (and generate filaments) whereas overexpression causes the production of too many septa, and hence minicells. The ZipA protein may be important for the localization of FtsZ because its N terminus is membrane-associated and its C terminus interacts with FtsZ. It is unclear how the septum is positioned in the cell, although nucleation sites probably exist because filaments resulting from temperature-sensitive *ftsZ* mutations rapidly form contractile rings at regular intervals when shifted to the permissive temperature. Genes of the *minB* locus limit septation to the central annulus and suppress the process at the terminal annuli remaining from previous cell divisions. *MinC* and *MinD* are septation inhibitors, whereas *MinE* antagonizes *MinCD* activity. The correct balance of all three

products thus inhibits terminal septation but protects the central annulus from inhibition.

The partition of bacterial chromosomes proceeds similarly to *plasmid partition* (q.v.). Partitioning may involve the association of the chromosome with the cell membrane, and may be regulated by replication: the origin and terminus of replication, as well as active replication forks, are membrane associated. Bacterial mutants which affect partition fall into two categories: those which interfere with the separation of interlocked replicated chromosomes (these include topoisomerase and *Xer* site-specific recombinase mutants) and those which affect the partition process itself. In the latter category no cis-acting sites have been found on the chromosome, but several trans-acting factors have been identified (e.g. the membrane protein *MukA* and the microtubule-associated protein *MukB*; mutations in both genes generate anucleate cells).

2.1 The eukaryotic cell cycle

The standard eukaryotic cell cycle.

The standard eukaryotic cell cycle is divided into four non-overlapping phases. The discrete events of the chromosome cycle (DNA synthesis and mitosis) occur during the ***S*** phase and the **M** phase, respectively, and in most cell cycles these landmarks are separated by **G1** and ***G2*** gap phases, during which mRNAs and proteins accumulate continuously (Figure 2.1).

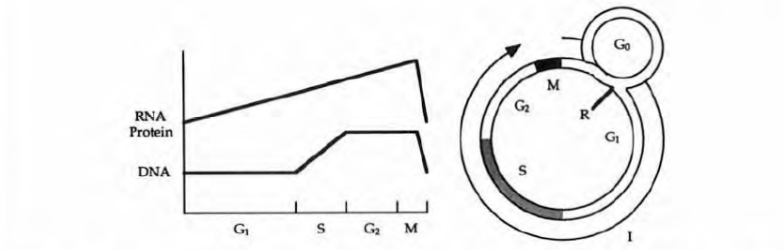


Figure 2.1: The standard eukaryotic cell cycle. The chromosome cycle is divided into four stages: G_1, **S** and G_2 which constitute the **interphase** (I), and **M** which is **mitosis.** The left panel shows typical relative durations of the cell cycle stages, although this varies in different species and depending on cell type and growth conditions. Animal cells can withdraw to the quiescent state, G_0, if growth factors are withdrawn in early G_1, but once the cell passes the **restriction point** (R) it becomes committed to a further round of DNA replication and division. The graph compares the accumulation of 'continuous' components with the discontinuous synthesis of DNA. The quantity of all cell components is halved at the end of the M phase when cell division occurs.

The process of crossing from one phase of the cell cycle to the next is a cell cycle transition. Whereas mitosis is a dramatic event that involves visible reorganization of cell structure, the rest of the cell cycle is unremarkable to the eye and is termed the interphase.

Variations on the theme (Table 2.2) include cell cycles where one or both gap phases are omitted, or where either the S phase or the M phase is omitted, leading to halving or doubling of the DNA content, respectively. In addition, a cell may be arrested (indefinitely or permanently delayed) at any stage of the cell cycle, as occurs during oocyte maturation and in postmitotic cells such as neurons. Animal cells may withdraw from the cell cycle altogether, entering a quiescent state termed *G_0*, where both growth and division are repressed. This reflects a continuing requirement for growth factors and other signaling molecules in the environment, and imposes an extra level of regulation on the cell cycle so that the growth and division of individual cells can be coordinated in the context of a multicellular organism. The abnormal cell proliferation seen in cancer is caused by the failure of this regulatory mechanism.

Table 2.1: Variations on the theme of the four-stage eukaryotic cell cycle

Modification	Circumstances
Stages omitted	
No gap phases	Rapid alternation between M and S phases is characteristic of early development in animals with large eggs because there is enough material in the egg for rapid cleavage divisions without cell growth. Many organisms miss out one or other of the gap phases: *Dictyostelium discoideum* replicates its DNA immediately following mitosis (no G_1), whereas *S. cerevisiae* undergoes mitosis directly after DNA replication (no G_2)
No S phase	Two rounds of division without intervening DNA synthesis occur during *meiosis* (q.v.)
No M phase	Multiple rounds of DNA synthesis without cell division occur in *Drosophila* secretory tissues to produce *polytene chromosomes* (q.v.)
Stages extended	
Indefinite arrest at G_1, G_2 or M	Oocytes and eggs may be arrested at G_1, G_2 or M depending on species. Fertilization releases the block and allows the cycle to resume
Withdrawal from the cell cycle at G_1 (G_0)	Many animal cells can withdraw from the cell cycle at G_1 and enter a quiescent state, often termed G_0, which may last months or years. This occurs if essential growth factors are withheld during early G_1 and involves the disassembly of the cell cycle control mechanism. Quiescent cells can be persuaded to re-enter the cycle if growth factors are made available, but there is a long delay before the initiation of the S phase while regulatory components are resynthesized. Normal somatic cells often enter G_0 after a characteristic number of divisions (the **Hayflick limit**), a phenomenon termed **senescence** which may be related to *telomere* length (q.v.) in some animals, and can also be induced by certain plasmids in fungi. Some cells withdraw entirely from the cell cycle as part of their differentiation and become **postmitotic**, e.g. neurons and muscle cells

Cell cycle check points:

The primary function of the cell cycle is to duplicate the genome precisely and divide it equally between two daughter cells. For this reason, it is important that the events of the cell cycle proceed in the correct order, and that each stage of the cell cycle is complete before the next commences. DNA content remains constant only if DNA replication alternates with mitosis, if mitosis occurs after the *completion* of DNA replication, and replication commences after mitosis has precisely divided the DNA. The cell meets these criteria by organizing the cell cycle as a dependent series of events. Thus, if mitosis is blocked, the cell cycle arrests at the M phase until the block is removed - it does not go ahead and replicate the DNA anyway (i.e. DNA replication is dependent upon the completion of mitosis). Similarly, if DNA replication is prevented, the cell does not attempt to undergo mitosis, because mitosis is dependent upon the completion of DNA replication.

A further function of the cell cycle is to coordinate the chromosome cycle with cell growth, so there is no progressive loss or gain of cytoplasm, and no cell proliferation in an unsuitable environment. Progress through the cell cycle is thus also dependent upon cell size and is regulated by nutrient availability, the presence of mating pheromones (in yeast), and the presence of growth factors and hormones (in animals).

The cell possesses a number of regulatory systems which can sense the progress of the cell cycle and can inhibit subsequent stages in the event of failure. These regulatory mechanisms are termed cell cycle checkpoints, and represent intrinsic signaling systems of cell cycle control. The checkpoint mechanisms also respond to external signals so that arrest may occur in cases of nutrient deprivation or growth factor withdrawal. There are numerous checkpoints in the cell cycle, which are clustered in two major groups - those occurring at G1 and regulating entry into the S phase, and those occurring at G2 and regulating entry into the M phase *(Figure* 2.2).

This clustering suggests that intrinsic and extrinsic signals may funnel into common components of cell cycle regulation. Additional checkpoints insure the orderly and dependent series of events which comprise mitosis. Different organisms attach varying degrees of significance to the G1 and G2 checkpoints, reflecting the stage at which the cell receives input from the environment. The G1 checkpoint is predominant in the budding yeast *Saccharomyces cereuisiae* (where it is called START) and in animal cells (where it is called the restriction point or commitment point). The yeast assesses nutrient availability and the presence of mating pheromones during G1, whereas animal cells respond to the presence of growth factors. Cells of both kingdoms will arrest at this checkpoint if the environment is unsuitable for growth, but once past it, they are committed to a round of DNA replication and mitosis regardless of their environment.

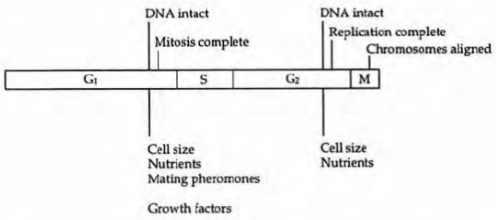


Figure 2.2: Known checkpoints in the eukaryotic cell cycle. These represent the points at which specific protein kinases are activated/inactivated.

Conversely, in the fission yeast *Schizosaccharomyces pombe,* the environment is monitored at the G2 checkpoint, and under satisfactory conditions the cell will undergo mitosis, division and the next round of DNA replication before checking again. The advantage of pausing at G2 rather than G1 for the haploid yeast cells reflects the presence of two copies of the genome at GP, allowing any damage to DNA to be repaired by recombination.

Studying cell cycle regulation

Two complementary approaches have been used to characterize and isolate the regulatory components of the cell cycle. In the heterokaryon approach, nuclei at different stages of the chromosome cycle are joined in a common cytoplasm and their behavior observed. Cultured mammalian cells and amphibian eggs have been used for these experiments. The results of fusing cultured fibroblasts synchronized at different cell cycle stages are shown in *Table* 2.2.

Table 2.2: Heterokaryon experiments to investigate regulatory factors controlling the cell cycle

Fusion	Result	Conclusion
$S \times G_1$	Both nuclei replicate	S nucleus contains an S-phase promoting factor
$S \times G_2$	S-phase cell completes replication, G_2-phase nucleus waits for S-phase nucleus to complete replication and then both cells enter the M phase	G_2 nucleus cannot respond to S-phase activator (a re-replication block), S-phase activator is also an inhibitor of mitosis
$M \times G_1$, S or G_2	Interphase nucleus enters precocious mitosis (regardless of state of chromosome replication)	M nucleus contains an M-phase promoting factor
$G_1 \times G_2$	Neither nucleus undergoes replication or mitosis	Both S-phase and M-phase activators are present transiently

The ability of M-phase cells to induce mitosis in any interphase nucleus provided early evidence for the existence of an M-phase promoting factor. Similar results were obtained in *Xenopus* nuclear transplantation studies - interphase nuclei formed spindles when injected into eggs arrested at the metaphase of meiosis I, and cytoplasm from these eggs could induce meiosis in oocytes arrested at G2. The large size of the eggs was exploited to purify the substance responsible, which was called maturation promoting factor. Further studies showed that maturation promoting factor could also induce mitosis in somatic cells, and was in fact identical to M-phase promoting factor, which shared the same acronym (MPF).

The second approach has been to exploit the versatility of yeast genetics to isolate conditional mutants for cell cycle

functions. Numerous ***cdc*** mutants (cell division cycle) have been identified which are blocked at various stages of the cell cycle, yet continue to grow. A second class of so-called ***wee*** mutants allows precocious transition of cell size checkpoints, and are smaller than wild-type cells. Many of the genes identified from ***cdc*** mutants are not specific cell cycle regulators, but control processes such **as** DNA replication, repair and mating, upon which the progress of the cell cycle depends. However, a number of ***cdc*** genes appear to play a direct role in the regulation of the cell cycle, as discussed in the following section. Satisfyingly, the biochemical analysis of MPF has shown that both approaches have converged on the same small group of molecules.

2.2. The molecular basis of cell cycle regulation

Cyclins and cyclin-dependent kinases (CDK)

The sequential stages of the cell cycle reflect alternative states of phosphorylation for key proteins which mediate the different cell cycle events. The cell cycle transitions represent switches in those phosphorylation states. The G1→S transition involves the phosphorylation of proteins required for DNA replication, whilst the G2→ M transition involves the phosphorylation of proteins required for mitosis. The basis of cell cycle regulation is a family of protein kinases which phosphorylate these target proteins and hence coordinate the different activities required for each transition.

The involvement of protein kinases in cell cycle control was revealed when analysis of S. *cerevisiue cdc* mutants blocked at START identified the product of the *CDC28* gene, a *34* kD protein kinase, as the principal regulator of the G1→S transition. The cdc2 gene, which played an equally important role in the G2→M transition in S. *pombe,* was found to encode a homologous protein kinase. Genes encoding similar kinases were subsequently isolated from vertebrates, and these could restore wild-type cell cycle function to yeast cdc mutants. Significantly, the *Xenopus* homolog of CDC28/Cdc2 was found to be a component of MPF.

The kinases were found to be present constitutively in the nucleus, but to control cell cycle transitions their activity would have to oscillate. An explanation for their periodic activity came from the study of sea urchin eggs, wherein were discovered a family of molecules whose synthesis and activity oscillated with the cell cycle. These molecules were termed cyclins and they were subsequently found in many other eukaryotes including yeast and vertebrates. The second component of MPF was found to be a B-type cyclin.

The activity of MPF resides in the catalytic kinase subunit but it is dependent upon the cyclin subunit, which introduces a conformational change in its partner to stimulate kinase activity. The cell cycle kinases are thus described as cyclin-dependent kinases (CDKs) and function as CDKcyclin holoenzymes. This strategy of cell cycle regulation appears to be conserved throughout the eukaryotes.

CDK-cyclin diversity in the yeast and animal cell cycles.

A number of potential CDK have been isolated from yeast, but only CDC28 in S. ***cerevisiue*** and Cdc2 in ***S. pombe*** appear to be directly involved in the cell cycle, and are required for the GI-S and the GYM transitions in both species. In animal cells, there is a greater diversity of CDKs. The first to be discovered, the p34 CDC28/Cdc2 component of MPF, appears to function specifically at the G2-M transition. Ten or more further CDKs are present in animal cells; five of these are involved specifically in the early stages of the cell cycle.

The diversity of cyclins is greater than that of CDKs, as different cyclins are synthesized at different stages of the cell cycle **in** both animals and yeast. There are at least eight families of vertebrate cyclins (designated A-H). Since CDKs phosphorylate different targets at each cell cycle transition, cyclins are required not only for kinase activity, but also for substrate specificity. In animals, alternative cyclins may be differentially expressed in different cell types, which would facilitate the unique aspects of cell cycle control in distinct differentiated cells. There are generally three types of cyclin in all organisms: the **G1** cyclins which regulate the G→S transition, **the S-phase** cyclins which are required for DNA replication, and the **M-phase** cyclins which are required for mitosis. M-phase cyclins include the S. ***cerevisiae*** CLB cyclins, the vertebrate A- and B-type cyclins and the ***S. pombe*** cyclin Cig 13. They are stable proteins but share a conserved motif called a destruction box, which **is** required for targeted ***ubiquitination*** resulting in degradation during mitosis.

Other cyclins are inherently unstable because they carry a ***PEST domain***, and their levels are determined primarily by the transcriptional activity of their genes. All cyclins carry a conserved motif, the cyclin box, which is required for CDK binding (Figure 2.3). The yeast and animal CDKs and cyclins involved in cell cycle regulation are summarized in ***Table*** 2.3.

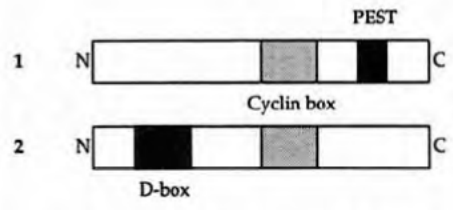


Figure 2.3: Domain structure of cyclins. (1) Cyclins which possess a PEST motif are targeted for proteolytic degradation and are very unstable. This class includes the *S. cerevisiae* CLN cyclins and vertebrate cyclins of classes C, D, E and F. Most of these are G_1/S cyclins. (2) Mitotic cyclins tend to be stable throughout interphase, but contain a destruction box required for their ubiquitin-dependent degradation during M phase. The first cyclins were isolated on the basis of their oscillating activity, but several are known to be synthesized constitutively and are defined on the basis of cyclin box homology rather than expression parameters.

Table 2.3: Principle CDKs and cyclins active at each stage of the yeast and mammalian cell cycles

Stage	*S. cerevisiae*	*S. pombe*	Mammals
G_1	CDK: CDC28 Cyclins: CLN1–3	CDK: Cdc2 Cyclins: Cig2	CDK: CDK2, 4, 5, 6 Cyclins: D1–3 CDK: CDK2 Cyclins: E Class
S-phase	CDK: CDC28 Cyclins: CLN5, CLN6	CDK: Cdc2 Cyclins: Cig2 ?	CDK: CDK2 Cyclins: A Class
G_2/M-phase	CDK: CDC28 Cyclins: CLB1–4	CDK: Cdc2 Cyclins: Cdc13	CDK: CDK1 (CDC2) Cyclins: A and B

Regulation of CDK-cyclin activity:

Cell cycle transitions are characterized by bursts of CDK activity which cause sudden switches in the phosphorylation states of target proteins responsible for cell cycle events. Sudden spikes of kinase activity are not regulated by cyclin synthesis and degradation alone, as cyclins accumulate gradually in the cell and only the mitotic cyclins are degraded by rapid, targeted proteolysis.

Two CDK-cyclin systems are active in G1 of the mammalian cell cycle. The CDK2-cyclin E complex is required for the GI-S transition. The other CDKs and the D cyclins are responsible for interpreting growth factor signals for the environment, and act athe restriction point to channel the cell into either late G1 or *Go.* The mammalian CDK1-cyclin B complex is MPF -the vertebrate homolog of yeast Cdc2/CDC28 may be termed Cdc2 or CDKl. CDK7/H cyclin, which is CAK, the mammalian CDK Thr-l61 kinase, is thought to be expressed constitutively.

CDKs phosphorylate target proteins, but are themselves also regulated by phosphorylation. The yeast CDC28 and Cdc2 CDKs are phosphorylated on two key residues, Tyr-15 and Thr-161. Phosphorylated Thr-161 is required for kinase activity, whereas phosphorylated Tyr-15 is inhibitory and dominant to Thr-161 phosphorylation. The principle determinant of CDK cyclin activity in yeast is thus the state of phosphorylation of Tyr-15, and some of the upstream regulatory components have been identified.

In S. ***pornbe,*** Wee1 is a tyrosine kinase which phosphorylates Cdc2 at Tyr-15 and thus inactivates it. Wee1 activity is antagonized by Cdc25 phosphatase, which removes phosphate groups from the same substrate. Both Wee1 and Cdc25 are themselves regulated by intrinsic and extrinsic signals, and this is believed to be the basis of the G2→M checkpoint in this species. The decision to proceed with mitosis or arrest in G2 thus reflects the relative levels of these opposing activities, and the regulatory networks which feed into this checkpoint are considered below.

Homologs of *wee2* and cdc25 have been identified in mammals, although the situation is more complex than in yeast because there are multiple isoforms which may demonstrate specificities for particular CDK-cyclin complexes. Additionally, there are three phosphorylation sites on mammalian CDC2 (CDK1), Thr-l4 and Tyr-15, both of which are dominant inhibitors when phosphorylated, and Thr-161, whose phosphorylation is required for kinase activity.

An enzyme has been identified in mammals which is responsible for Thr-161 phosphorylation. Remarkably, this turns out to be yet another CDK-cyclin complex comprising CDK7 and cyclin H; it is known as CDK-activating kinase (CAK). The enzyme responsible for Thr-14 phosophorylation has not been identified. CDK-cyclin complexes are also regulated by inhibitory proteins. This third level of control is used for both intrinsic and extrinsic regulation purposes. As discussed below, the S. *pornbe* Rum1 protein is a specific inhibitor of the mitotic CDK-cyclin complex, and is synthesized throughout the G1 and S phases, thus preventing the cycle skipping DNA replication and entering mitosis prematurely.

The FAR1 protein is activated in response to signaling by mating-type pheromones in S. *cerevisiae* and inhibits the START CDK-cyclin complex, thus arresting the cell at G1 in readiness for mating. Two families of CDK-cyclin inhibitors (CKIs) are found in animals. One blocks all CDK-cyclin activities and the other specifically inhibits D-cyclin complexes containing CDK4 and CDK6 (whose major substrate is the retinoblastoma protein). These regulatory mechanisms are discussed in more detail below.

2.3 Progress through the cell cycle

Transition of the START checkpoint in yeast

START is the predominant checkpoint in the S. *cerevisiae* cell cycle and is well characterized in this species *(Figure* 2.4). In GI, the CDC28 CDK can associate with any of three cyclins (CLN1, CLN2, CLN3) which are functionally redundant, at least under laboratory conditions (the genes can be deleted singly or in pairs with no effect on phenotype, but mutation in all three arrests the cell cycle at GI). The highly unstable CLN3 is expressed constitutively and acts as an indicator of cell growth. When the concentration of CDC28-CLN3 reaches a critical level, the kinase activates two transcription factors, SBF and MBF, which contain a common component, SWI6. SBF initiates transcription of the genes for the other G1 cyclins, CLNl and CLN2.

These cyclins associate with CDC28 to form CDK-cyclin complexes with several functions related to the activation of the six CLB cyclins (the mitotic cyclins CLB1, CLBZ, CLB3 and CLB4, and the S-phase cyclins CLB5 and CLBG), which share a conserved destruction box. Firstly, they reduce the rate of proteolysis of the CLB cyclins, and secondly, they increase the rate of hydrolysis of a CDK-CLB cyclin inhibitor, SIC1.

The second transcription factor, MBF, initiates transcription of the *CLBS* and *CLB6* genes. Thus, by stimulating the transcription of Sphase cyclins and increasing their stability, the G1-specific CDK-cyclin complexes induce the onset of the S phase when cell growth is sufficient. The S-phasespecific CDK-cyclin complexes stimulate the transcription or activity of replication proteins, perhaps including those found at the constitutive origin recognition complexes on yeast ARS elements.

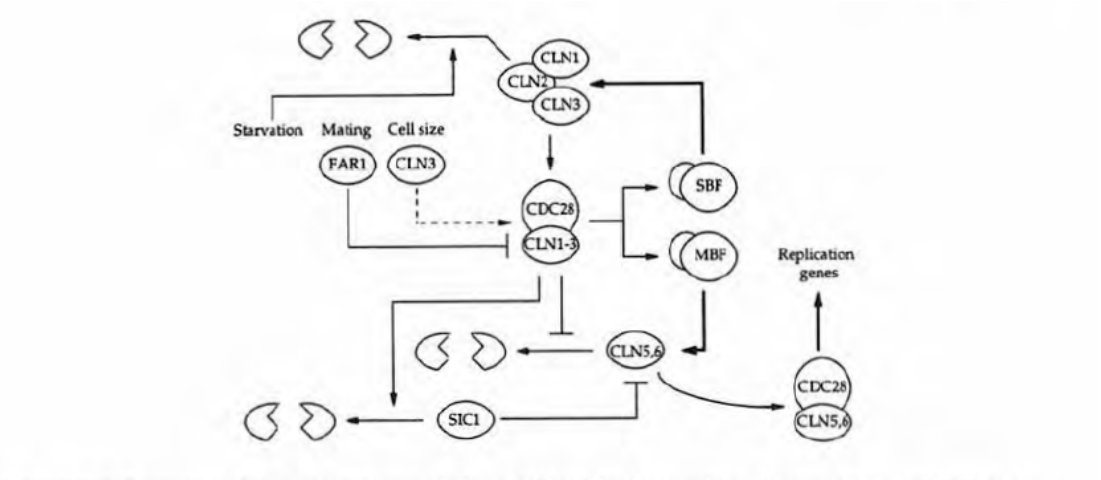


Figure 2.4: Summary of the molecular regulation of START in *S. cerevisiae*. Accumulation of CLN3 acts as an indicator of cell size (dashed line). At a critical concentration, the kinase activity of CDC28 is induced and activates two transcription factors. SBF promotes *CLN1–3* transcription, thus generating a positive feedback loop, although proteolytic degradation of CLN2 due to starvation can delay entry into the S phase. MBF activates the genes for CLN5 and CLN6 whilst activated CDC28 kinase inhibits destruction box-targeted cyclin degradation and promotes degradation of the inhibitor SIC1. CLN5 and CLN6 form complexes with CDC28 which activate the expression of replication genes.

The cell cannot pass START until mitosis is complete and in S. *pombe* this is achieved by a simple negative feedback mechanism in which the mitotic CDK-cyclin complex inhibits a protein required for the initiation of DNA replication, Cdc18. In the absence of active Cdc18, DNA replication is blocked, and this repression is lifted when the mitotic kinase is destroyed.

Entry into the S phase in mammalian cells

In mammals, four CDKs (CDK2, CDK3, CDK4 and CDK6) act in early GI. All four associate with the D cyclins, which are synthesized early in G1 in a growth-factordependent manner. CDK4 and CDK6 play the principle roles in the regulation of downstream events. The only known target of these early CDK-cyclin complexes is the retinoblastoma protein RB-1, which is a negative regulator of the cell cycle. RB-1 in its unphosphorylated form binds transcription factors of the E2F family which normally activate genes required for entry into S phase. RB-1 inhibits the expression of these genes in two ways: by sequestering the E2F activation domain, and direct repression by chromatin remodeling. Phosphorylation of RB-1 by the early G1 CDK-D cyclin complexes releases the repression. An important downstream effect of CDK-D cyclin activation is the synthesis of E cyclins, which are required for the GI-S transition itself. The E cyclins form complexes with CDK2, and these activate CDC25A phosphatase, which may in turn activate the S phase-specific complex CDK2/A cyclin. This complex is required for the initiation of replication, and has been localized at replication origins where it can be immunoprecipitated with *PCNA*, a component of the major eukaryotic DNA polymerase. Its specific targets in the initiation complex remain to be characterized, but both cyclin A and cyclin E complexes are known to phosphorylate and regulate the activity of several transcription factors including the E2F family, p53, B-Myb and the inhibitory helix-loop-helix protein Id2.

The restriction of replication to once per cell cycle

The cell fusion studies described above show that the S-phase nucleus can induce the G1 nucleus to replicate but not the ***G2*** or M nuclei, which need to complete mitosis before becoming competent for replication once again. In all cases, replication occurs only once. A model proposed to explain these data involves a replication licensing factor displaying the following properties:

- it interacts with DNA origins in the latter stages of mitosis;
- it is essential for initiation of replication, but is inactivated or destroyed once initiation has taken place;
- it is unable to cross the nuclear envelope.

The model proposes that a cytoplasmic licensing factor interacts with DNA origins before reassembly of the nuclear envelope at the telophase. This factor is a target for regulation by S-phase promoting CDK-cyclin complexes and initiates replication at each origin at the beginning of the S phase, becoming inactive in the process. During the remainder of the cell cycle, the new licensing factor is unable to enter the nucleus, and the precocious initiation of replication is prevented. The factor enters the nucleus at the start of mitosis, but is inactive or prevented from interacting with DNA because of the condensed state of the chromatin. The factor interacts with origins as the
matin decondenses, but remains inactive, awaiting the S-phase signal. The nuclear envelope closes, preventing entry by further licensing factor, and any uncomplexed factor in the nucleus, and perhaps also in the cytoplasm, is targeted for degradation.
Proteins which display many of the properties of the licensing factor have been identified in *Xenopus,* mammals and yeast. In *S. cerevisiae,* the product of the *CDC46* gene is thought to be the licensing factor, or a component of it, although in yeast the nuclear envelope does not break down during mitosis and the licensing factor is transported into the nucleus by translocation.

Control of entry into mitosis

In yeast, entry into mitosis is controlled by the same cyclin dependent kinase that regulates START (CDC2, CDC28), but in combination with specific mitotic cyclins. In ***S. cerevisiae,*** these are encoded by the ***CLB1, CLB2, CLB3*** and ***CLB4*** genes, and in ***S. pombe*** by the ***cdc13***
gene. **In** vertebrates, the CDKl(CDC2) kinase associates with A- and B-type cyclins which, like their yeast counterparts, contain mitotic destruction boxes. Much of the regulatory

pathway of the G2-M checkpoint has been determined in S. pombe, and similar components are found in other eukaryotes. Essentially, mitosis is initiated by a burst of Cdc2 activity generated by a posttranslational positive feedback mechanism (this differs from entry into the S phase in ***S. cerevisiae,*** which is governed by a transcriptional feedback loop). The mechanism is broadly conserved throughout the eukaryotes, but the checkpoints for entry into mitosis are regulated in different ways, as discussed below.

In ***S. pombe,*** Cdc2 cyclin-dependent kinase appears to be regulated predominantly by phosphorylation of the Tyr-15 residue ***(Figure 2.5).***

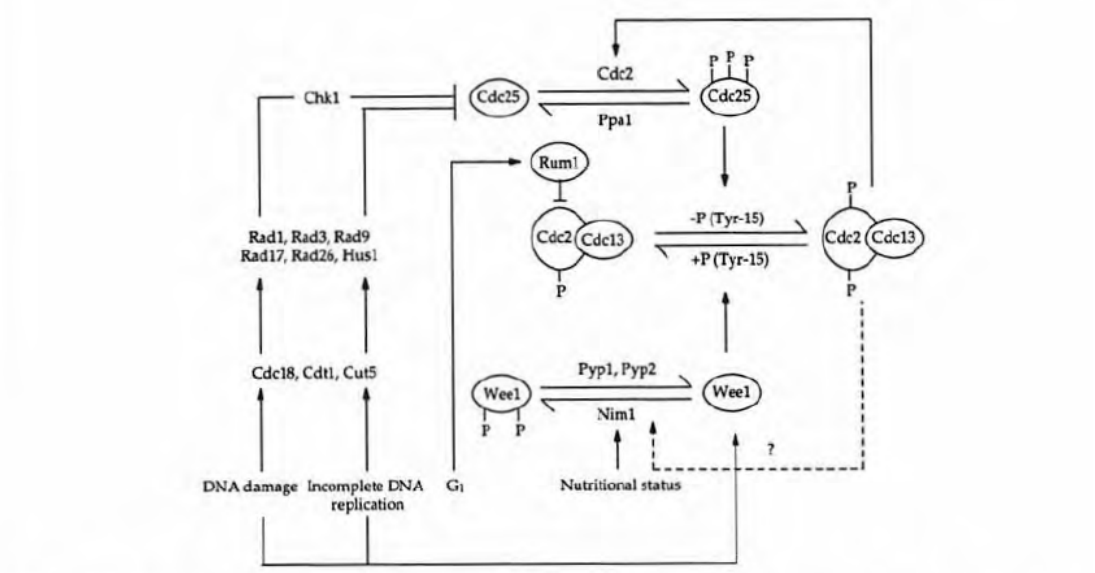


Figure 2.5: Control of entry into mitosis in *S. pombe.* Entry into mitosis requires active Cdc2/Cdc13 kinase which phosphorylates and activates a number of proteins required for the appropriate structural changes in the cell. Cdc2/Cdc13 activity is controlled predominantly by phosphorylation of Cdc2 Tyr-15, which is regulated by the opposing activities of Cdc25 phosphatase and Wee1 kinase. Both these regulators are themselves regulated by intrinsic and extrinsic signals (e.g. incomplete DNA replication, nutritional status), as discussed in the text.

It is activated by dephosphorylation, carried out by Cdc25 and another phosphatase termed Pyp3, and inactivated by redundant kinases termed Weel and Mikl. The G2-M transition coincides with a burst of Cdc2 kinase activity which is controlled by a positive feedback loop in which the Cdc2/Cdcl3 kinase phosphorylates and activates Cdc25, which in turn dephosphorylates and activates Cdc2 (Cdc2 may also inactivate Weel). This sudden exponential rise in kinase activity simultaneously changes the phosphorylation states of many target proteins and effectively throws a switch which initiates mitosis ***(Table 2.4).***

Table 2.4: Structural changes in the cell at the beginning of the M phase and the role of M-phase kinase

Substrates	Role in mitosis
Cdc25	Positive feedback loop to generate surge of M-phase kinase activity
Anaphase promoting complex (This is likely to be an indirect substrate)	Negative feedback loop to induce degradation of cyclin. Causes transient spike of M-phase kinase activity at onset of the M phase
Histone H1, HMG proteins	Condensation of chromatin
TFIID, TFIIIB, poly(A) polymerase	Inhibition of basal transcription
SBF, SWI5 (*S. cerevisiae*)	Regulation of transcription (phosphorylation of SWI5 inhibits nuclear import)
Lamins	Phosphorylation of nuclear lamins causes dissociation into subunits which breaks down nuclear lamina and may drive dissolution of the nuclear membrane
RMSA (regulator of mitotic spindle assembly)	Mitotic spindle assembly
Vimentin	Reorganization of cytoskeleton
Caldesmon, Myosin light chain	Caldesmon is an actin filament regulator and MLC interacts with actin to form contractile ring at the site of cytokinesis. Phosphorylation inhibits this activity
Caesein kinase II α and β subunits, c-Src, c-Abl	Protein kinases with regulatory roles or specificity for other downstream targets?

Note that the nuclear membrane breaks up in higher eukaryotes but not in yeast, where the nucleus as well as the cell undergoes division (**karyokinesis**).

Wee1 and Cdc25 process intrinsic signals and integrate signals from the environment, and are probably the main checkpoint regulators. Wee1 activation is regulated by a kinase Niml, which is responsive to the nutritional status of the cell. Checks for completion of DNA replication or the presence of damaged DNA are mediated through a group of genes, including *cdcl8, cdtl* and *cut5,* whose products are thought to form a recognition complex which is monitored by Radl, Rad3, Rad9, Radl7, Rad26 and Husl. The downstream targets for these signals may be Cdc25 or Weel or both. Rad3 has recently been shown to phosphorylate Chkl, which in turn phosphorylates Cdc25 and causes it to be sequestered in a complex with 14-3-3 protein. In animals, there may also be a mechanism which regulates CDC2 kinase activity independently of phosphorylation on Tyr-15.

The checkpoints for DNA damage and DNA replication are different in *S. cerevisiae,* which lacks a definite G2 phase. In this species, the cell forms a spindle and arrests in metaphase in response to DNA damage or incomplete replication. The target for the checkpoint control is the anaphase promoting complex (APC), which is thought to promote the metaphase-anaphase transition by linking ubiquitin to the mitotic cyclin destruction box. The DNA damage checkpoint appears to be mediated by a protein called PDS1, which may control sister chromatid separation or may function indirectly by inhibiting APC. Interestingly, the DNA replication checkpoint of *Aspergillus niduluns* involves both control of Cdc2 (through Wee1 and Cdc25) and a downstream component, NimA kinase, through the product of the *bimE* gene. BimE is a component of the *A. niduluns* APC.

Mitosis:

The M-phase CDK-cyclin complex phosphorylates target proteins mediating the structural changes occurring during mitosis (condensation of chromatin, assembly of the spindle apparatus, reorganization of the nucleus and assembly of the cytokinetic machinery), although few targets have been precisely defined *(Table 2.4).* One critical but probably indirect target of the M-phase kinase is the anaphase-promoting complex (APC), a destruction box-dependent ubiquitin ligase which facilitates the sudden degradation of mitotic cyclins (in animals, cyclin A at metaphase and cyclin B at the metaphase-anaphase transition - the differential timing may reflect the efficiencies of the destruction boxes). Other APC targets include proteins required for maintaining the association of sister chromatids, possibly kinetochore components or perhaps a more generally distributed factor - a candidate is S. *cerevisiue* PDS1, as discussed above. Another important target of M-phase kinase is the myosin light chain, which in its phopshorylated state is prevented from binding to actin and thus from forming the cytoskeletal ring required for cytokinesis. Degradation of cyclin B is dependent upon chromatin alignment on the mitotic spindle, so it is possible that formation of the APC is dependent upon kinetochore attachment and inhibited by unattached kinetochores. The inactivation of M-phase kinase (triggered by degradation of cyclin B) allows M-phase kinase substrates such as myosin to be dephosphorylated, thus facilitating the cytoskeletal organization which precludes cytokinesis, the reconstruction of the nuclear lamina and decondensation of chromatin. The cell ensures the correct order of events in mitosis, as for the cell cycle as a whole, by organizing them into a dependent series *(Table* 2.5).

Table 2.5: The stages of mitosis with important cellular events indicated and molecular mechanisms (where known) in italics

Stage	Critical events
Prophase	*M-phase kinase activated.* Chromatin condenses and mitotic spindle begins to form. These are probably direct results of phosphorylation of target proteins by M-phase kinase
Prometaphase	In metazoans, breakdown of nuclear envelope allows mitotic spindle access to chromosomes. Some spindle microtubules attach to kinetochores. Nuclear reorganization is also directly regulated by M-phase kinase. When all chromosomes attached to spindle, *APC activated by CDC20/Slp1 following inhibition of MAD/BUB pathway*
Metaphase	*Cyclin A degraded.* Chromosomes guided to **metaphase plate** held by tension between opposing poles of spindle, which are attached to sister kinetochores
Anaphase	*Cyclin B degraded; M-phase kinase inactive.* Paired kinetochores separate towards poles as kinetochore microtubules shorten; poles also move apart as polar microtubules repel each other. *Proteolytic machinery responsible for cyclin B degradation probably also targets a protein which binds sister chromatids together (Scc1/Mcd1). This stage must be dependent on completed chromosome alignment. The nature of the early metaphase signal which prevents cyclin B destruction and precocious segregation is unknown, but constitutes a major cell cycle checkpoint*
Telophase	*Substrates of M-phase kinase dephosphorylated (including myosin light chain, nuclear lamins, histones).* Chromosomes reach poles, nuclear membrane reforms, chromatin decondenses, contractile ring assembles
Cytokinesis	Contractile ring completed. Cell constricts around remains of mitotic spindle and cleaves into two daughter cells

2.4 Special cell cycle systems in animals

Exit from the cell cycle

As discussed above, the withdrawal of growth factors from animal cells at a critical period during G1 causes them to cease growth and exit from the cell cycle, entering a quiescent state termed Go. This may be a transient response to the withdrawal of growth factors, a permanent aspect of differentiation (e.g. in neurons and muscle cells), or regulated during development to control the final size of a growing structure. The absence of growth factors stimulates exit from the cell cycle because growth factors are required for the synthesis of the D cyclins; these are very unstable proteins replenished only by new transcription early in GI. It is this transcription which is activated by growth factor

signaling. In summary, growth factors are ligands for receptor tyrosine kinases which initiate a signal transduction cascade through Ras, Raf and MAP kinases, eventually activating transcription factors in the nucleus. The targets for these transcription factors are a set of so-called immediate early genes, some of which also encode transcriptional regulators likely to activate D-cyclin transcription. Go cells stimulated with growth factors enter the S phase several hours later, reflecting the time taken for D-cyclin mRNA to appear in the cytoplasm.

Withdrawal from the cell cycle can also be stimulated by growth inhibitors such as TGF-P, contact inhibition or loss of substrate. These diverse signals act through a family of small cyclin dependent kinase inhibitors (CKIs) targeting specific components of the cell cycle machinery. The p16 family of inhibitors specifically inhibit CDK-D cyclins by preventing both their assembly and their ability to phosphorylate the RB-1 protein (*tumor-suppressor genes).* The p21 and p27 families inhibit all CDK-cyclin activities and thus arrest the cell cycle by inhibiting the activity of CDK-cyclin D complexes by preventing their activation by CAK (which is itself a CDK-cyclin holoenzyme) and also by blocking the activity of cyclin E and cyclin A-containing complexes, and the PCNA/6-DNA polymerase complex itself. The CKIs. also integrate signals from intrinsic pathways. A major pathway of DNA damage regulation works through the p53 regulator which activates p21.

As well as stimulating D-cyclin synthesis, growth factors facilitate cell growth by repressing many of the CDKs discussed above. Ironically, it has been found that p21 or p27 are constitutively present in the cell and comprise part of the normal, functional CDK-cyclin complex. Stoichiometric binding of the inhibitory proteins does not affect kinase activity, but inhibition occurs at higher concentrations.
p21 is also capable of direct interaction with PCNA/DNA polymerase δ, so that DNA damage blocks DNA replication directly rather than signaling through a CDK-dependent kinase cascade.

Apoptosis as an alternative to cell cycle arrest.

Apoptosis is programed cell death, a form of cell death initiated by the cell itself in response to various intrinsic and external signals (as opposed to necrosis which is caused by damage, infection or injury). Apoptosis has an important role in development and is also a defence against chemical insult and infection. The inhibition of apoptosis can lead to cancer.

The behavior of a cell undergoing apoptosis is distinct from that of a necrotic cell. It involves condensation and peripheralization of chromatin, loss of cytoplasm, fragmentation of the nucleus, compaction of organelles, the disestablishment of communication with neighboring cells, fusion of the endoplasmic reticulum with the outer cell membrane and finally fragmentation of the cell into numerous apoptotic bodies which are engulfed by surrounding cells. There is no inflammation as seen with necrosis. Prior to chromatin condensation, DNA is degraded into small fragments representing multiples of the nucleosome unit.

Central to the control of apoptosis is the BCL-2 family of cell death regulators. BCL-2 itself is a survival factor (i.e. it inhibits apoptosis) and is homologous to the C. *elegans* CED-9 protein expressed in all surviving cells during development. Other family members (e.g. BM, BAD and BAK) are stimulators of apoptosis. The proteins form dimers, and competition between them, reflecting their relative abundances, determines the fate of the cell. The effectors of BCL-2 signaling are a family of cysteine proteases related to the interleukin-lp converting enzyme (ICE proteases).

A number of such enzymes are synthesized by mammalian cells as inactive zymogens and are activated in a proteolytic cascade, culminating in the activation of proteins such as poly(ADP)ribose polymerase and nuclear lamins, the direct effectors of apoptotic cell behavior.

The signal transduction pathways leading to apoptosis are not fully understood. One pathway involves p53, which initiates apoptosis in response to DNA damage in some cell types (e.g. skin cells) by inhibiting BCL-2 and activating BAX. In other cells, p53 arrests the cell cycle in G1 by activating the CKI proteins which inhibit CDK-cyclin complexes and *stress activated* protein *kinases*. This pathway appears not to be involved in apoptosis occurring during development (e.g. in the interdigital regions of the mouse limb bud) as mice homozygous for a deletion of *TP53* develop normally.

Other signals, e.g. growth inhibitory factors such as TGF-P, can also induce apoptosis, and do so through BCL-2 family proteins but independently of p53. There are also signaling pathways influencing the ICE family proteases directly. For example, tumor necrosis factor (TNF) initiates the apoptotic signaling cascade by revealing an 80 residue cytoplasmic death domain which is shared by the TNF receptor and several othereceptors and intracellular apoptotic signaling molecules (e.g. TRADD, Fas, MORT1 and RIP), all of which activate the ICE proteases. CrmA is an ICE protease inhibitor encoded by cowpox virus. This is a survival factor which blocks apoptosis induced by p53, TGF-P and TNF, allowing the virus to avoid the lethal consequences of this response to infection.

E. Microbial Physiology: Growth, yield and characteristics, strategies of cell division, stress response.

Growth may be defined as an increase in cellular constituents. It leads to a rise in cell number when microorganisms reproduce by processes like budding or binary fission. In the latter, individual cells enlarge and divide to yield two progeny of approximately equal size.

Growth also results when cells simply become longer or larger. If the microorganism is coenocytic—that is, a multinucleate organism in which nuclear divisions are not accompanied by cell divisions— growth results in an increase in cell size but not cell number. It is usually not convenient to investigate the growth and reproduction of individual microorganisms because of their small size. Therefore, when studying growth, microbiologists normally follow changes in the total population number.

1 The Growth Curve

Population growth is studied by analyzing the growth curve of a microbial culture. When microorganisms are cultivated in liquid medium, they usually are grown in a batch culture or closed system— that is, they are incubated in a closed culture vessel with a single batch of medium. Because no fresh medium is provided during incubation, nutrient concentrations decline and concentrations of wastes increase. The growth of microorganisms reproducing by binary fission can be plotted as the logarithm of the number of viable cells versus the incubation time. The resulting curve has four distinct phases (figure 1).

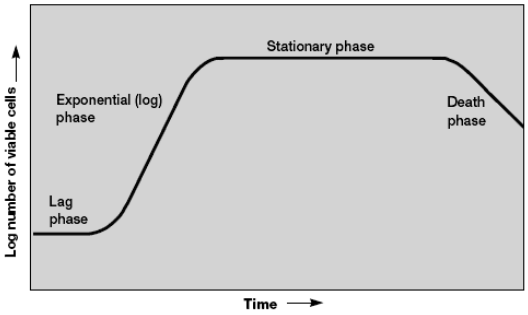


Figure 1 Microbial Growth Curve in a Closed System.

Lag Phase

When microorganisms are introduced into fresh culture medium, usually no immediate increase in cell number occurs, and therefore this period is called the lag phase. Although cell division does not take place right away and there is no net increase in mass, the cell is synthesizing new components. A lag phase prior to the start of cell division can be necessary for a variety of reasons. The cells may be old and depleted of ATP, essential cofactors, and ribosomes; these must be synthesized before growth can begin. The medium may be different from the one the microorganism was growing in previously. Here new enzymes would be needed to use different nutrients. Possibly the microorganisms have been injured and require time to recover. Whatever the causes, eventually the cells retool, replicate their DNA, begin to increase in mass, and finally divide.

The lag phase varies considerably in length with the condition of the microorganisms and the nature of the medium. This phase may be quite long if the inoculum is from an old culture or one that has been refrigerated. Inoculation of a culture into a chemically different medium also results in a longer lag phase. On the other hand, when a young, vigorously growing exponential phase culture is transferred to fresh medium of the same composition, the lag phase will be short or absent.

Exponential Phase

During the exponential or log phase, microorganisms are growing and dividing at the maximal rate possible given their genetic potential, the nature of the medium, and the conditions under which they are growing. Their rate of growth is constant during the exponential phase; that is, the microorganisms are dividing and doubling in number at regular intervals. Because each individual divides at a slightly different moment, the growth curve rises smoothly rather than in discrete jumps (figure 1). The population is most uniform in terms of chemical and physiological properties during this phase; therefore exponential phase cultures are usually used in biochemical and physiological studies.

Exponential growth is balanced growth. That is, all cellular constituents are manufactured at constant rates relative to each other. If nutrient levels or other environmental conditions change, unbalanced growth results. This is growth during which the rates of synthesis of cell components vary relative to one another until a new balanced state is reached. This response is readily observed in a shift-up experiment in which bacteria are transferred from a nutritionally poor medium to a richer one. The cells first construct new ribosomes to enhance their capacity for protein synthesis. This is followed by increases in protein and DNA synthesis. Finally, the expected rise in reproductive rate takes place.

Unbalanced growth also results when a bacterial population is shifted down from a rich medium to a poor one. The organisms may previously have been able to obtain many cell components directly from the medium. When shifted to a nutritionally inadequate medium, they need time to make the enzymes required for the biosynthesis of unavailable nutrients. Consequently cell division and DNA replication continue after the shift-down, but net protein and RNA synthesis slow. The cells become smaller and reorganize themselves metabolically until they are able to grow again. Then balanced growth is resumed and the culture enters the exponential phase.

These shift-up and shift-down experiments demonstrate that microbial growth is under precise, coordinated control and responds quickly to changes in environmental conditions. When microbial growth is limited by the low concentration of a required nutrient, the final net growth or yield of cells increases with the initial amount of the limiting nutrient present (figure 2a). This is the basis of microbiological assays for vitamins and other growth factors. The rate of growth also increases with nutrient concentration (figure 2b), but in a hyperbolic manner much like that seen with many enzymes.

The shape of the curve seems to reflect the rate of nutrient uptake by microbial transport proteins. At sufficiently high nutrient levels the transport systems are saturated, and the growth rate does not rise further with increasing nutrient concentration.

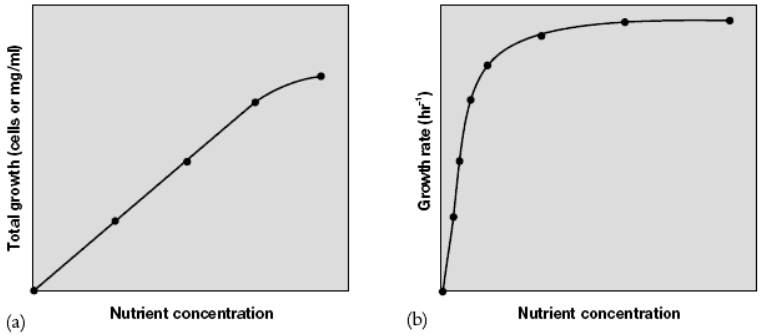


Figure 2 Nutrient Concentration and Growth.

(a) The effect of changes in limiting nutrient concentration on total microbial yield. At sufficiently high concentrations, total growth will plateau.
(b) The effect on growth rate.

Stationary Phase

Eventually population growth ceases and the growth curve becomes horizontal (figure 1). This stationary phase usually is attained by bacteria at a population level of around 10^9 cells per ml. Other microorganisms normally do not reach such high population densities, protozoan and algal cultures often having maximum concentrations of about 10^6 cells per ml. Of course final population size depends on nutrient availability and other factors, as well as the type of microorganism being cultured. In the stationary phase the total number of viable microorganisms remains constant. This may result from a balance between cell division and cell death, or the population may simply cease to divide though remaining metabolically active. Microbial populations enter the stationary phase for several reasons. One obvious factor is nutrient limitation; if an essential
nutrient is severely depleted, population growth will slow. Aerobic organisms often are limited by O_2 availability. Oxygen is not very soluble and may be depleted so quickly that only the surface of a culture will have an O_2 concentration adequate for growth.

The cells beneath the surface will not be able to grow unless the culture is shaken or aerated in another way. Population growth also may cease due to the accumulation of toxic waste products. This factor seems to limit the growth of many anaerobic cultures (cultures growing in the absence of O_2). For example, streptococci can produce so much lactic acid and other organic acids from sugar fermentation that their medium becomes acidic and growth is inhibited. Streptococcal cultures also can enter the stationary phase due to depletion of their sugar supply. Finally, there is some evidence that growth may cease when a critical population level is reached. Thus entrance into the stationary phase may result from several factors operating in concert.

As we have seen, bacteria in a batch culture may enter stationary phase in response to starvation. This probably often occurs in nature as well because many environments have quite low nutrient levels. Starvation can be a positive experience for bacteria. Many do not respond with obvious morphological changes such as endospore formation, but only decrease somewhat in overall size, often accompanied by protoplast shrinkage and nucleoid condensation. The more important changes are in gene expression and physiology. Starving bacteria frequently produce a variety of starvation proteins, which make the cell much more resistant to damage in a variety of ways. They increase peptidoglycan cross-linking and cell wall strength. The Dps (DNA-binding protein from starved cells) protein protects DNA. Chaperones prevent protein denaturation and renature damaged proteins. As a result of these and many other mechanisms, the starved cells become harder to kill and more resistant to starvation itself, damaging temperature changes, oxidative and osmotic damage, and toxic chemicals such as chlorine. These changes are so effective that some bacteria can survive starvation for years. Clearly, these considerations are of great practical importance in medical and industrial microbiology. There is even evidence that Salmonella typhimurium and some other bacterial pathogens become more virulent when starved.

Death Phase

Detrimental environmental changes like nutrient deprivation and the buildup of toxic wastes lead to the decline in the number of viable cells characteristic of the death phase. The death of a microbial population, like its growth during the exponential phase, is usually logarithmic (that is, a constant proportion of cells dies every hour). This pattern in viable cell count holds even when the total cell number remains constant because the cells simply fail to lyse after dying. Often the only way of deciding whether a bacterial cell is viable is by incubating it in fresh medium; if it does not grow and reproduce, it is assumed to be dead. That is, death is defined to be the irreversible loss of the ability to reproduce. Although most of a microbial population usually dies in a logarithmic fashion, the death rate may decrease after the population has been drastically reduced. This is due to the extended survival of particularly resistant cells. For this and other reasons, the death phase curve may be complex.

The Mathematics of Growth

Knowledge of microbial growth rates during the exponential phase is indispensable to microbiologists. Growth rate studies contribute to basic physiological and ecological research and the solution of applied problems in industry. Therefore the quantitative aspects of exponential phase growth will be discussed. During the exponential phase each microorganism is dividing at constant intervals. Thus the population will double in number during a specific length of time called the generation time or doubling time. This situation can be illustrated with a simple example.

Suppose that a culture tube is inoculated with one cell that divides every 20 minutes (table 1). The population will be 2 cells after 20 minutes, 4 cells after 40 minutes, and so forth. Because the population is doubling every generation, the increase in population is always 2n where n is the number of generations. The resulting population increase is exponential or logarithmic (figure 3).

Table 1 An Example of Exponential Growth

Time[a]	Division Number	2^n	Population ($N_0 \times 2^n$)	$\log_{10} N_t$
0	0	$2^0 = 1$	1	0.000
20	1	$2^1 = 2$	2	0.301
40	2	$2^2 = 4$	4	0.602
60	3	$2^3 = 8$	8	0.903
80	4	$2^4 = 16$	16	1.204
100	5	$2^5 = 32$	32	1.505
120	6	$2^6 = 64$	64	1.806

[a]The hypothetical culture begins with one cell having a 20-minute generation time.

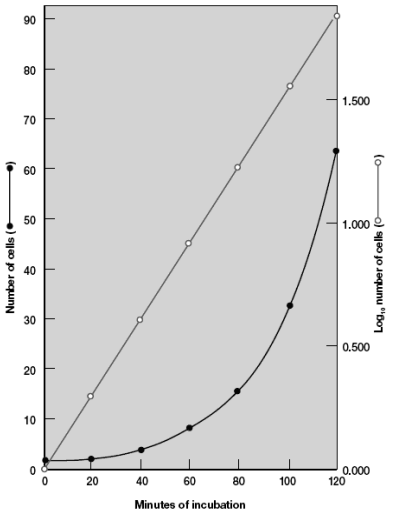


Figure 3 **Exponential Microbial Growth.** The data from table 1 for six generations of growth are plotted directly (●—●) and in the logarithmic form (○—○). The growth curve is exponential as shown by the linearity of the log plot.

These observations can be expressed as equations for the generation time.

Let N_0 = the initial population number
N_t = the population at time t
n = the number of generations in time t

Then inspection of the results in table 1 will show that

$N_t = N_0 \times 2^n$.

Solving for n, the number of generations, where all logarithms are to the base 10,

$$\log N_t = \log N_0 + n \cdot \log 2, \text{ and}$$

$$n = \frac{\log N_t - \log N_0}{\log 2} = \frac{\log N_t - \log N_0}{0.301}$$

The rate of growth during the exponential phase in a batch culture can be expressed in terms of the mean growth rate constant (k).

This is the number of generations per unit time, often expressed as the generations per hour.

$$k = \frac{n}{t} = \frac{\log N_t - \log N_0}{0.301t}$$

The time it takes a population to double in size—that is, the mean generation time or mean doubling time (g), can now be calculated. If the population doubles (t = g), then

$N_t = 2 N_0$.

Substitute $2N_0$ into the mean growth rate equation and solve for k.

$$k = \frac{\log (2N_0) - \log N_0}{0.301g} = \frac{\log 2 + \log N_0 - \log N_0}{0.301g}$$

$$k = \frac{1}{g}$$

The mean generation time is the reciprocal of the mean growth rate constant.

$$g = \frac{1}{k}$$

The mean generation time (g) can be determined directly from a semilogarithmic plot of the growth data (figure 4) and the growth rate constant calculated from the g value. The generation time also may be calculated directly from the previous equations. For example, suppose that a bacterial population increases from 10^3 cells to 10^9 cells in 10 hours.

$$k = \frac{\log 10^9 - \log 10^3}{(0.301)(10 \text{ hr})} = \frac{9 - 3}{3.01 \text{ hr}} = 2.0 \text{ generations/hr}$$

$$g = \frac{1}{2.0 \text{ gen./hr}} = 0.5 \text{ hr/gen. or } 30 \text{ min/gen.}$$

Generation times vary markedly with the species of microorganism and environmental conditions. They range from less than 10 minutes (0.17 hours) for a few bacteria to several days with some eucaryotic microorganisms (table 2). Generation times in nature are usually much longer than in culture.

2 The Influence of Environmental Factors on Growth

Microorganisms must be able to respond to variations in nutrient levels, and particularly to nutrient limitation. The growth of microorganisms also is greatly affected by the chemical and physical nature of their surroundings. An understanding of environmental influences aids in the control of microbial growth and the study of the ecological distribution of microorganisms.

The ability of some microorganisms to adapt to extreme and inhospitable environments is truly remarkable. Procaryotes are present anywhere life can exist. Many habitats in which prokaryotes thrive would kill most other organisms. Procaryotes such as Bacillus infernus even seem able to live over 1.5 miles below the Earth's surface, without oxygen and at temperatures above 60°C. Microorganisms that grow in such harsh conditions are often called extremophiles.

Table 2 Generation Times for Selected Microorganisms

Microorganism	Temperature (°C)	Generation Time (Hours)
Bacteria		
Beneckea natriegens	37	0.16
Escherichia coli	40	0.35
Bacillus subtilis	40	0.43
Staphylococcus aureus	37	0.47
Pseudomonas aeruginosa	37	0.58
Clostridium botulinum	37	0.58
Rhodospirillum rubrum	25	4.6–5.3
Anabaena cylindrica	25	10.6
Mycobacterium tuberculosis	37	≈12
Treponema pallidum	37	33
Algae		
Scenedesmus quadricauda	25	5.9
Chlorella pyrenoidosa	25	7.75
Asterionella formosa	20	9.6
Euglena gracilis	25	10.9
Ceratium tripos	20	82.8
Protozoa		
Tetrahymena geleii	24	2.2–4.2
Leishmania donovani	26	10–12
Paramecium caudatum	26	10.4
Acanthamoeba castellanii	30	11–12
Giardia lamblia	37	18
Fungi		
Saccharomyces cerevisiae	30	2
Monilinia fructicola	25	30

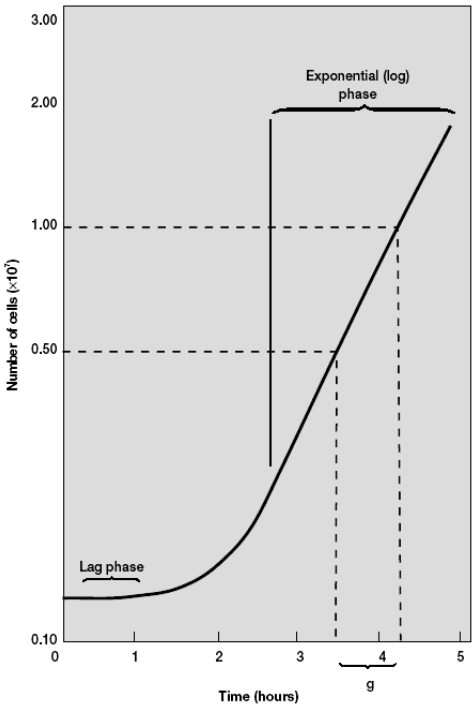


Figure 4 Generation Time Determination. The generation time can be determined from a microbial growth curve. The population data are plotted with the logarithmic axis used for the number of cells. The time to double the population number is then read directly from the plot. The log of the population number can also be plotted against time on regular axes.

In this section we shall briefly review how some of the most important environmental factors affect microbial growth. Major emphasis will be given to solutes and water activity, pH, temperature, oxygen level, pressure, and radiation. Table 3 summarizes the way in which microorganisms are categorized in terms of their response to these factors.

Solutes and Water Activity

Because a selectively permeable plasma membrane separates microorganisms from their environment, they can be affected by changes in the osmotic concentration of their surroundings. If a microorganism is placed in a hypotonic solution (one with a lower osmotic concentration), water will enter the cell and cause it to burst unless something is done to prevent the influx. The osmotic concentration of the cytoplasm can be reduced by use of inclusion bodies. Procaryotes also can contain pressure-sensitive channels that open to allow solute escape when the osmolarity of the environment becomes much lower than that of the cytoplasm. Most bacteria, algae, and fungi have rigid cell walls that maintain the shape and integrity of the cell. When microorganisms with rigid cell walls are placed in a hypertonic environment, water leaves and the plasma membrane shrinks away from the wall, a process known as plasmolysis. This dehydrates the cell and may damage the plasma membrane; the cell usually becomes metabolically inactive and ceases to grow. Many microorganisms keep the osmotic concentration of their protoplasm somewhat above that of the habitat by the use of compatible solutes, so that the plasma membrane is always pressed firmly against their cell wall. Compatible solutes are solutes that are compatible with metabolism and growth when at high intracellular concentrations. Most procaryotes increase their internal osmotic concentration in a hypertonic environment through the synthesis or uptake of choline, betaine, proline, glutamic acid, and other amino acids; elevated levels of potassium ions are also involved to some extent. Algae and fungi employ sucrose and polyols—for example, arabitol, glycerol, and mannitol— for the same purpose. Polyols and amino acids are ideal solutes for this function because they normally do not disrupt enzyme structure and function. A few procaryotes like Halobacterium salinarium raise their osmotic concentration with potassium ions (sodium ions are also elevated but not as much as potassium). Halobacterium's enzymes have been altered so that they actually require high salt concentrations for normal activity. Since protozoa do not have a cell wall, they must use contractile vacuoles to eliminate excess water when living in hypotonic environments.

Table 3 Microbial Responses to Environmental Factors

Descriptive Term	Definition	Representative Microorganisms
Solute and Water Activity		
Osmotolerant	Able to grow over wide ranges of water activity or osmotic concentration	*Staphylococcus aureus, Saccharomyces rouxii*
Halophile	Requires high levels of sodium chloride, usually above about 0.2 M, to grow	*Halobacterium, Dunaliella, Ectothiorhodospira*
pH		
Acidophile	Growth optimum between pH 0 and 5.5	*Sulfolobus, Picrophilus, Ferroplasma, Acontium, Cyanidium caldarium*
Neutrophile	Growth optimum between pH 5.5 and 8.0	*Escherichia, Euglena, Paramecium*
Alkalophile	Growth optimum between pH 8.5 and 11.5	*Bacillus alcalophilus, Natronobacterium*
Temperature		
Psychrophile	Grows well at 0°C and has an optimum growth temperature of 15°C or lower	*Bacillus psychrophilus, Chlamydomonas nivalis*
Psychrotroph	Can grow at 0–7°C; has an optimum between 20 and 30°C and a maximum around 35°C	*Listeria monocytogenes, Pseudomonas fluorescens*
Mesophile	Has growth optimum around 20–45°C	*Escherichia coli, Neisseria gonorrhoeae, Trichomonas vaginalis*
Thermophile	Can grow at 55°C or higher; optimum often between 55 and 65°C	*Bacillus stearothermophilus, Thermus aquaticus, Cyanidium caldarium, Chaetomium thermophile*
Hyperthermophile	Has an optimum between 80 and about 113°C	*Sulfolobus, Pyrococcus, Pyrodictium*
Oxygen Concentration		
Obligate aerobe	Completely dependent on atmospheric O_2 for growth.	*Micrococcus luteus, Pseudomonas, Mycobacterium;* most algae, fungi, and protozoa
Facultative anaerobe	Does not require O_2 for growth, but grows better in its presence.	*Escherichia, Enterococcus, Saccharomyces cerevisiae*
Aerotolerant anaerobe	Grows equally well in presence or absence of O_2	*Streptococcus pyogenes*
Obligate anaerobe	Does not tolerate O_2 and dies in its presence.	*Clostridium, Bacteroides, Methanobacterium, Trepomonas agilis*
Microaerophile	Requires O_2 levels below 2–10% for growth and is damaged by atmospheric O_2 (20%).	*Campylobacter, Spirillum volutans, Treponema pallidum*
Pressure		
Barophilic	Growth more rapid at high hydrostatic pressures.	*Photobacterium profundum, Shewanella benthica, Methanococcus jannaschii*

The amount of water available to microorganisms can be reduced by interaction with solute molecules (the osmotic effect) or by adsorption to the surfaces of solids (the matric effect). Because the osmotic concentration of a habitat has such profound effects on microorganisms, it is useful to be able to express quantitatively the degree of water availability. Microbiologists generally use water activity (a_w) for this purpose (water availability also may be expressed as water potential, which is related to a_w). The water activity of a solution is 1/100 the relative humidity of the solution (when expressed as a percent). It is also equivalent to the ratio of the solution's vapor pressure (P_{soln}) to that of pure water (P_{water}).

$$a_w = \frac{P_{soln}}{P_{water}}$$

The water activity of a solution or solid can be determined by sealing it in a chamber and measuring the relative humidity after the system has come to equilibrium. Suppose after a sample is treated in this way, the air above it is 95% saturated—that is, the air contains 95% of the moisture it would have when equilibrated at the same temperature with a sample of pure water. The relative humidity would be 95% and the sample's water activity, 0.95. Water activity is inversely related to osmotic pressure; if a solution has high osmotic pressure, its aw is low.

Microorganisms differ greatly in their ability to adapt to habitats with low water activity (table 4). A microorganism must expend extra effort to grow in a habitat with a low aw value because it must maintain a high internal solute concentration to retain water. Some microorganisms can do this and are osmotolerant; they will grow over wide ranges of water activity or osmotic concentration. For example,

Staphylococcus aureus can be cultured in media containing any sodium chloride concentration up to about 3 M. It is well adapted for growth on the skin. The yeast Saccharomyces rouxii will grow in sugar solutions with a_w values as low as 0.6. The alga Dunaliella viridis tolerates sodium chloride concentrations from 1.7 M to a saturated solution. Although a few microorganisms are truly osmotolerant, most only grow well at water activities around 0.98 (the approximate aw for seawater) or higher. This is why drying food or adding large quantities of salt and sugar is so effective in preventing food spoilage. As table 4 shows, many fungi are osmotolerant and thus particularly important in the spoilage of salted or dried foods.

Table .4 Approximate Lower a_w Limits for Microbial Growth

Water Activity	Environment	Bacteria	Fungi	Algae
1.00—Pure water	Blood, Plant wilt, Seawater { Vegetables, meat, fruit	Most gram-negative nonhalophiles		
0.95	Bread	Most gram-positive rods	*Basidiomycetes*	Most algae
0.90	Ham	Most cocci, *Bacillus*	*Fusarium* *Mucor, Rhizopus* Ascomycetous yeasts	
0.85	Salami	*Staphylococcus*	*Saccharomyces rouxii* (in salt)	
0.80	Preserves		*Penicillium*	
0.75	Salt lakes Salted fish	*Halobacterium* *Actinospora*	*Aspergillus*	*Dunaliella*
0.70	Cereals, candy, dried fruit		*Aspergillus*	
0.60	Chocolate Honey Dried milk		*Saccharomyces rouxii* (in sugars) *Xeromyces bisporus*	
0.55—DNA disordered				

Halophiles have adapted so completely to hypertonic, saline conditions that they require high levels of sodium chloride to grow, concentrations between about 2.8 M and saturation (about 6.2 M) for extreme halophilic bacteria. The archaeon Halobacterium can be isolated from the Dead Sea (a salt lake between Israel and Jordan and the lowest lake in the world), the Great Salt Lake in Utah, and other aquatic habitats with salt concentrations approaching saturation. Halobacterium and other extremely halophilic bacteria have significantly modified the structure of their proteins and membranes rather than simply increasing the intracellular concentrations of solutes, the approach used by most osmotolerant microorganisms. These extreme halophiles accumulate enormous quantities of potassium in order to remain hypertonic to their environment; the internal potassium concentration may reach 4 to 7 M. The enzymes, ribosomes, and transport proteins of these bacteria require high levels of potassium for stability and activity. In addition, the plasma membrane and cell wall of Halobacterium are stabilized by high concentrations of sodium ion. If the sodium concentration decreases too much, the wall and plasma membrane literally disintegrate. Extreme halophilic bacteria have successfully adapted to environmental conditions that would destroy most organisms. In the process they have become so specialized that they have lost ecological flexibility and can prosper only in a few extreme habitats.

pH

pH is a measure of the hydrogen ion activity of a solution and is defined as the negative logarithm of the hydrogen ion concentration (expressed in terms of molarity).

$$pH = -\log [H^+] = \log(1/[H^+])$$

The pH scale extends from pH 0.0 (1.0 M H^+) to pH 14.0 (1.0×10^{14} M H^+), and each pH unit represents a tenfold change in hydrogen ion concentration.

It is not surprising that pH dramatically affects microbial growth. Each species has a definite pH growth range and pH growth optimum. Acidophiles have their growth optimum between pH 0 and 5.5; neutrophiles, between pH 5.5 and 8.0; and alkalophiles prefer the pH range of 8.5 to 11.5. Extreme alkalophiles have growth optima at pH 10 or higher. In general, different microbial groups have characteristic pH preferences. Most bacteria and protozoa are neutrophiles. Most fungi prefer slightly acid surroundings, about pH 4 to 6; algae also seem to favor slight acidity. There are many exceptions to these generalizations. For example, the alga Cyanidium caldarium and the archaeon Sulfolobus acidocaldarius are common inhabitants of acidic hot springs; both grow well around pH 1 to 3 and at high temperatures. The Archaea Ferroplasma acidarmanus and Picrophilus oshimae can actually grow at pH 0, or very close to it. Although microorganisms will often grow over wide ranges of pH and far from their optima, there are limits to their tolerance. Drastic variations in cytoplasmic pH can harm microorganisms by disrupting the plasma membrane or inhibiting the activity of enzymes and membrane transport proteins. Procaryotes die if the internal pH drops much below 5.0 to 5.5. Changes in the external pH also might alter the ionization of nutrient molecules and thus reduce their availability to the organism.

Several mechanisms for the maintenance of a neutral cytoplasmic pH have been proposed. The plasma membrane may be relatively impermeable to protons. Neutrophiles appear to exchange potassium for protons using an antiport transport system. Extreme alkalophiles like Bacillus alcalophilus maintain their internal pH closer to neutrality by exchanging internal sodium ions for external protons. Internal buffering also may contribute to pH homeostasis.

Microorganisms often must adapt to environmental pH changes to survive. In bacteria, potassium/proton and sodium/proton antiport systems probably correct small variations in pH. If the pH becomes too acidic, other mechanisms come into play. When the pH drops below about 5.5 to 6.0, Salmonella typhimurium and E. coli synthesize an array of new proteins as part of what has been called their acidic tolerance response. A proton-translocating ATPase contributes to this protective response, either by making more ATP or by pumping protons out of the cell. If the external pH decreases to 4.5 or lower, chaperones such as acid shock proteins and heat

shock proteins are synthesized. Presumably these prevent the acid denaturation of proteins and aid in the refolding of denatured proteins. Microorganisms frequently change the pH of their own habitat by producing acidic or basic metabolic waste products. Fermentative microorganisms form organic acids from carbohydrates, whereas chemolithotrophs like Thiobacillus oxidize reduced sulfur components to sulfuric acid. Other microorganisms make their environment more alkaline by generating ammoniathrough amino acid degradation.

Buffers often are included in media to prevent growth inhibition by large pH changes. Phosphate is a commonly used buffer and a good example of buffering by a weak acid ($H_2PO_4^-$) and its conjugate base ($HPO4^{2-}$).

$$H^+ + HPO_4^{2-} \longrightarrow H_2PO_4^-$$

$$OH^- + H_2PO_4^- \longrightarrow HPO_4^{2-} + HOH$$

If protons are added to the mixture, they combine with the salt form to yield a weak acid. An increase in alkalinity is resisted because the weak acid will neutralize hydroxyl ions through proton donation to give water. Peptides and amino acids in complex media also have a strong buffering effect.

Temperature

Environmental temperature profoundly affects microorganisms, like all other organisms. Indeed, microorganisms are particularly susceptible because they are usually unicellular and their temperature varies with that of the external environment. For these reasons, microbial cell temperature directly reflects that of the cell's surroundings. A most important factor influencing the effect of temperature on growth is the temperature sensitivity of enzyme catalyzed reactions. At low temperatures a temperature rise increases the growth rate because the velocity of an enzyme-catalyzed reaction, like that of any chemical reaction, will roughly double for every 10°C rise in temperature. Because the rate of each reaction increases, metabolism as a whole is more active at higher temperatures, and the microorganism grows faster. Beyond a certain point further increases actually slow growth, and sufficiently high temperatures are lethal. High temperatures damage microorganisms by denaturing enzymes, transport carriers, and other proteins. Microbial membranes are also disrupted by temperature extremes; the lipid bilayer simply melts and disintegrates.

Thus, although functional enzymes operate more rapidly at higher temperatures, the microorganism may be damaged to such an extent that growth is inhibited because the damage cannot be repaired. At very low temperatures, membranes solidify and enzymes don't work rapidly. In summary, when organisms are above the optimum temperature, both function and cell structure are affected. If temperatures are very low, function is affected but not necessarily cell chemical composition and structure.

Because of these opposing temperature influences, microbial growth has a fairly characteristic temperature dependence with distinct cardinal temperatures—minimum, optimum, and maximum growth temperatures (figure 5). Although the shape of the temperature dependence curve can vary, the temperature optimum is always closer to the maximum than to the minimum.

The cardinal temperatures for a particular species are not rigidly fixed but often depend to some extent on other environmental factors such as pH and the available nutrients. For example, Crithidia fasciculata, a flagellated protozoan living in the gut of mosquitos, will grow in a simple medium at 22 to 27°C. However, it cannot be cultured at 33 to 34°C without the addition of extra metals, amino acids, vitamins, and lipids.

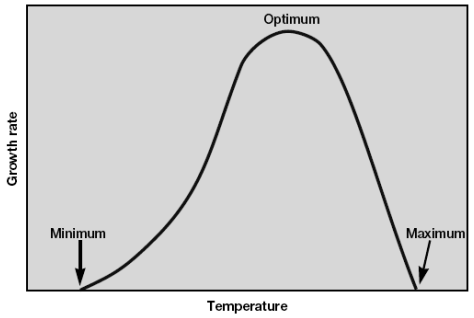


Figure 5

Temperature and Growth. The effect of temperature on growth rate.

The cardinal temperatures vary greatly between microorganisms. Optima normally range from 0°C to as high as 75°C, whereas microbial growth occurs at temperatures extending from -20°C to over 100°C. The major factor determining this growth range seems to be water. Even at the most extreme temperatures, microorganisms need liquid water to grow. The growth temperature range for a particular microorganism usually spans about 30 degrees. Some species (e.g., Neisseria gonorrhoeae) have a small range; others, like Enterococcus faecalis, will grow over a wide range of temperatures. The major microbial groups differ from one another regarding their maximum growth temperature. The upper limit for protozoa is around 50°C. Some algae and fungi can grow at temperatures as high as 55 to 60°C. Procaryotes have been found growing at or close to 100°C, the boiling point of water at sea level. Recently strains growing at even higher temperatures have been discovered. Clearly, prokaryotic organisms can grow at much higher temperatures than eucaryotes. It has been suggested that eucaryotes are not able to manufacture organellar membranes that are stable and functional at temperatures above 60°C. The photosynthetic apparatus also appears to be relatively unstable because photosynthetic organisms are not found growing at very high temperatures.

Microorganisms can be placed in one of five classes based on their temperature ranges for growth (figure 6.13).

1. **Psychrophiles** grow well at 0°C and have an optimum growth temperature of 15°C or lower; the maximum is around 20°C. They are readily isolated from Arctic and Antarctic habitats; because 90% of the ocean is 5°C or colder, it constitutes an enormous habitat for psychrophiles. The psychrophilic alga Chlamydomonas nivalis can actually turn a snowfield or glacier pink with its bright red spores. Psychrophiles are widespread among bacterial taxa and found in such genera as Pseudomonas, Vibrio, Alcaligenes, Bacillus, Arthrobacter, Moritella, Photobacterium, and Shewanella. The psychrophilic archaeon Methanogenium has recently been isolated from Ace Lake in Antarctica. Psychrophilic microorganisms have adapted to their environment in several ways. Their enzymes, transport systems, and protein synthetic mechanisms function well at low temperatures. The cell membranes of psychrophilic microorganisms have high levels of unsaturated fatty acids and remain semifluid when cold. Indeed, many psychrophiles begin to leak cellular constituents at temperatures higher than 20°C because of cell membrane disruption.

2. Many species can grow at 0 to 7°C even though they have optima between 20 and 30°C, and maxima at about 35°C. These are called **psychrotrophs or facultative psychrophiles.** Psychrotrophic bacteria and fungi are major factors in the spoilage of refrigerated foods.

3. **Mesophiles** are microorganisms with growth optima around 20 to 45°C; they often have a temperature minimum of 15 to 20°C. Their maximum is about 45°C or lower. Most microorganisms probably fall within this category. Almost all human pathogens are mesophiles, as might be expected since their environment is a fairly constant 37°C.

4. Some microorganisms are **thermophiles**; they can grow at temperatures of 55°C or higher. Their growth minimum is usually around 45°C and they often have optima between 55 and 65°C. The vast majority are procaryotes although a few algae and fungi are thermophilic. These organisms flourish in many habitats including composts, self-heating hay stacks, hot water lines, and hot springs. Thermophiles differ from mesophiles in having much more heat-stable enzymes and protein synthesis systems able to function at high temperatures. Their membrane lipids are also more saturated than those of mesophiles and have higher melting points; therefore thermophile membranes remain intact at higher temperatures.

5. As mentioned previously, a few thermophiles can grow at 90°C or above and some have maxima above 100°C. Procaryotes that have growth optima between 80°C and about 113°C are called **hyperthermophiles**. They usually do not grow well below 55°C. Pyrococcus abyssi and Pyrodictium occultum are examples of marine hyperthermophiles found in hot areas of the seafloor.

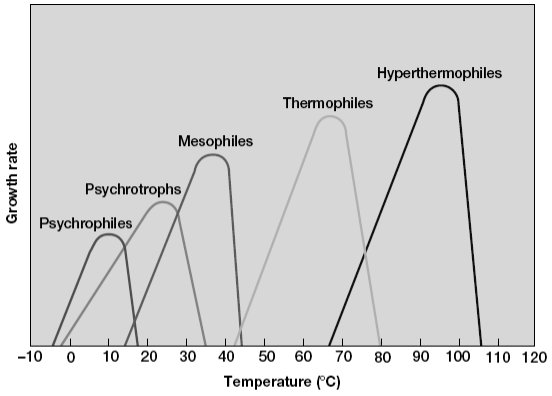


Figure 6. Temperature Ranges for Microbial Growth. Microorganisms can be placed in different classes based on their temperature ranges for growth. They are ranked in order of increasing growth temperature range as psychrophiles, psychrotrophs, mesophiles, thermophiles, and hyperthermophiles. Representative ranges and optima for these five types are illustrated here.

Oxygen Concentration

An organism able to grow in the presence of atmospheric O_2 is an aerobe, whereas one that can grow in its absence is an anaerobe. Almost all multicellular organisms are completely dependent on atmospheric O_2 for growth—that is, they are obligate aerobes. Oxygen serves as the terminal electron acceptor for the electron-transport chain in aerobic respiration. In addition, aerobic eukaryotes employ O_2 in the synthesis of sterols and unsaturated fatty acids.

- **Facultative anaerobes** do not require O_2 for growth but do grow better in its presence. In the presence of oxygen they will use aerobic respiration.
- **Aerotolerant anaerobes** such as Enterococcus faecalis simply ignore O_2 and grow equally well whether it is present or not.
- In contrast, strict or **obligate anaerobes** (e.g., Bacteroides, Fusobacterium, Clostridium pasteurianum, Methanococcus) do not tolerate O_2 at all and die in its presence. Aerotolerant and strict anaerobes cannot generate energy through respiration and must employ fermentation or anaerobic respiration pathways for this purpose.
- Finally, there are aerobes such as Campylobacter, called **microaerophiles**, that are damaged by the normal atmospheric level of O_2 (20%) and require O_2 levels below the range of 2 to 10% for growth. The nature of bacterial O_2 responses can be readily determined by growing the bacteria in culture tubes filled with a solid culture medium or a special medium like thioglycollate broth, which contains a reducing agent to lower O_2 levels

A microbial group may show more than one type of relationship to O_2. All five types are found among the procaryotes and protozoa. Fungi are normally aerobic, but a number of species— particularly among the yeasts—are facultative anaerobes. Algae are almost always obligate aerobes. It should be noted that the ability to grow in both aerobic and anaerobic environments provides considerable flexibility and is an ecological advantage. Although strict anaerobes are killed by O_2, they may be recovered from habitats that appear to be aerobic. In such cases they associate with facultative anaerobes that use up the available O_2 and thus make the growth of strict anaerobes possible. For example, the strict anaerobe Bacteroides gingivalis lives in the mouth where it grows in the anaerobic crevices around the teeth. These different relationships with O_2 appear due to several factors, including the inactivation of proteins and the effect of toxic O_2 derivatives. Enzymes can be inactivated when sensitive groups like sulfhydryls are oxidized. A notable example is the nitrogen-fixation enzyme nitrogenase, which is very oxygen sensitive.

Oxygen accepts electrons and is readily reduced because its two outer orbital electrons are unpaired. Flavoproteins, several other cell constituents, and radiation promote oxygen reduction. The result is usually some combination of the reduction products superoxide radical, hydrogen peroxide, and hydroxyl radical.

$$O_2 + e^- \rightarrow O_2^{-\cdot} \text{ (superoxide radical)}$$

$$O_2^{-\cdot} + e^- + 2H^+ \longrightarrow H_2O_2 \text{ (hydrogen peroxide)}$$

$$H_2O_2 + e^- + H^+ \longrightarrow H_2O + OH\cdot \text{ (hydroxyl radical)}$$

These products of oxygen reduction are extremely toxic because they are powerful oxidizing agents and rapidly destroy cellular constituents. A microorganism must be able to protect itself against such oxygen products or it will be killed. Neutrophils and macrophages use these toxic oxygen products to destroy invading pathogens.

Many microorganisms possess enzymes that afford protection against toxic O2 products. Obligate aerobes and facultative anaerobes usually contain the enzymes superoxide dismutase (SOD) and catalase, which catalyze the destruction of superoxide radical and hydrogen peroxide, respectively. Peroxidase also can be used to destroy hydrogen peroxide.

$$2O_2^{-\cdot} + 2H^+ \xrightarrow{\text{superoxide dismutase}} O_2 + H_2O_2$$

$$2H_2O_2 \xrightarrow{\text{catalase}} 2H_2O + O_2$$

$$H_2O_2 + NADH + H^+ \xrightarrow{\text{peroxidase}} 2H_2O + NAD^+$$

Aerotolerant microorganisms may lack catalase but almost always have superoxide dismutase. The aerotolerant Lactobacillus plantarum uses manganous ions instead of superoxide dismutase to destroy the superoxide radical. All strict anaerobes lack both enzymes or have them in very low concentrations and therefore cannot tolerate O_2.

Because aerobes need O_2 and anaerobes are killed by it, radically different approaches must be used when growing the two types of microorganisms. When large volumes of aerobic microorganisms are cultured, either the culture vessel is shaken to aerate the medium or sterile air must be pumped through the culture vessel. Precisely the opposite problem arises with anaerobes; all O_2 must be excluded. This can be accomplished in several ways. (1) Special anaerobic media containing reducing agents such as thioglycollate or cysteine may be used. The medium is boiled during preparation to dissolve its components; boiling also drives off oxygen very effectively. The reducing agents will eliminate any dissolved O2 remaining within the medium so that anaerobes can grow beneath its surface. (2) Oxygen also may be eliminated from an anaerobic system by removing air with a vacuum pump and flushing out residual O_2 with nitrogen gas. Often CO_2 as well as nitrogen is added to the chamber since many anaerobes require a small amount of CO_2 for best growth. (3) One of the most popular ways of culturing small numbers of anaerobes is by use of a Gas-Pak jar (figure 7). In this procedure the environment is made anaerobic by using hydrogen and a palladium catalyst to removeO2 through the formation of water. The reducing agents in anaerobic agar also remove oxygen, as mentioned previously. (4) Plastic bags or pouches make convenient containers when only a few samples are to be incubated anaerobically. These have a catalyst and calcium carbonate to produce an anaerobic, carbon-dioxiderich atmosphere. A special solution is added to the pouch's reagent compartment; petri dishes or other containers are placed in the pouch; it then is clamped shut and placed in an incubator.

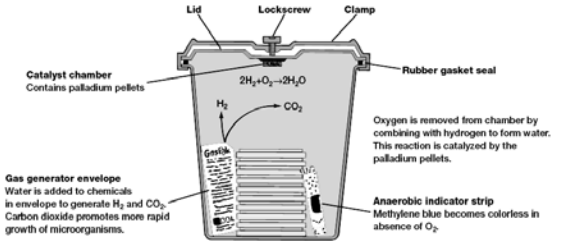


Figure 7 The GasPak Anaerobic System. Hydrogen and carbon dioxide are generated by a GasPak envelope. The palladium catalyst in the chamber lid catalyzes the formation of water from hydrogen and oxygen, thereby removing oxygen from the sealed chamber.

Pressure

Most organisms spend their lives on land or on the surface of water, always subjected to a pressure of 1 atmosphere (atm), and are never affected significantly by pressure. Yet the deep sea (ocean of 1,000 m or more in depth) is 75% of the total ocean volume. The hydrostatic pressure can reach 600 to 1,100 atm in the deep sea, while the temperature is about 2 to 3°C. Despite these extremes, bacteria survive and adapt. Many are barotolerant: increased pressure does adversely affect them but not as much as it does nontolerant bacteria. Some bacteria in the gut of deep-sea invertebrates such as amphipods and holothurians are truly barophilic—they grow more rapidly at high pressures. These gut bacteria may play an important role in nutrient recycling in the deep sea. One barophile has been recovered from the Mariana trench near the Philippines (depth about 10,500 m) that is actually unable to grow at pressures below about 400 to 500 atm when incubated at 2°C. Thus far, barophiles have been found among several bacterial genera (e.g., Photobacterium, Shewanella, Colwellia). Some members of the Archaea are thermobarophiles (e.g., Pyrococcus spp., Methanococcus jannaschii).

Radiation

Our world is bombarded with electromagnetic radiation of various types. This radiation often behaves as if it were composed of waves moving through space like waves traveling on the surface of water. The distance between two wave crests or troughs is the wavelength. As the wavelength of electromagnetic radiation decreases, the energy of the radiation increases—gamma rays and X rays are much more energetic than visible light or infrared waves. Electromagnetic radiation also acts like a stream of energy packets called photons, each photon having a quantum of energy whose value will depend on the wavelength of the radiation. Sunlight is the major source of radiation on the Earth. It includes visible light, ultraviolet (UV) radiation, infrared rays, and radio waves. Visible light is a most conspicuous and important aspect of our environment: all life is dependent on the ability of photosynthetic organisms to trap the light energy of the sun. Almost 60% of the sun's radiation is in the infrared region rather than the visible portion of the spectrum. Infrared is the major source of the Earth's heat. At sea level, one finds very little ultraviolet radiation below about 290 to 300 nm. UV radiation of wavelengths shorter than 287 nm is absorbed by O_2 in the Earth's atmosphere; this process forms a layer of ozone between 25 and 30 miles above the Earth's surface. The ozone layer then absorbs somewhat longer UV rays and reforms O_2. This elimination of UV radiation is crucial because it is quite damaging to living systems. The fairly even distribution of sunlight throughout the visible spectrum accounts for the fact that sunlight is generally "white."

Many forms of electromagnetic radiation are very harmful to microorganisms. This is particularly true of ionizing radiation, radiation of very short wavelength or high energy, which can cause atoms to lose electrons or ionize. Two major forms of ionizing radiation are (1) X rays, which are artificially produced, and (2) gamma rays, which are emitted during radioisotope decay. Low levels of ionizing radiation will produce mutations and may indirectly result in death, whereas higher levels are directly lethal. Although microorganisms are more resistant to ionizing radiation than larger organisms, they will still be destroyed by a sufficiently large dose. Ionizing radiation can be used to sterilize items. Some procaryotes (e.g., Deinococcus radiodurans) and bacterial endospores can survive large doses of ionizing radiation

A variety of changes in cells are due to ionizing radiation; it breaks hydrogen bonds, oxidizes double bonds, destroys ring structures, and polymerizes some molecules. Oxygen enhances these destructive effects, probably through the generation of hydroxyl radicals ($OH^{\cdot}$). Although many types of constituents can be affected, it is reasonable to suppose that destruction of DNA is the most important cause of death.

Ultraviolet (UV) radiation, mentioned earlier, kills all kinds of microorganisms due to its short wavelength (approximately from 10 to 400 nm) and high energy. The most lethal UV radiation has a wavelength of 260 nm, the wavelength most effectively absorbed by DNA. The primary mechanism of UV damage is the formation of thymine dimers in DNA. Two adjacent thymines in a DNA strand are covalently joined to inhibit DNA replication and function. This damage is repaired in several ways. In photoreactivation, blue light is used by a photoreactivating enzyme to split thymine dimers. A short sequence containing the thymine dimer can also be excised and

replaced. This process occurs in the absence of light and is called dark reactivation. Damage also can be repaired by the recA protein in recombination repair and SOS repair. When UV exposure is too heavy, the damage is so extensive that repair is impossible.

Although very little UV radiation below 290 to 300 nm reaches the earth's surface, near-UV radiation between 325 and 400 nm can harm microorganisms. Exposure to near-UV radiation induces tryptophan breakdown to toxic photoproducts. It appears that these toxic tryptophan photoproducts plus the near-UV radiation itself produce breaks in DNA strands. The precise mechanism is not known, although it is different from that seen with 260 nm UV. Visible light is immensely beneficial because it is the source of energy for photosynthesis. Yet even visible light, when present in sufficient intensity, can damage or kill microbial cells. Usually pigments called photosensitizers and O_2 are required. All microorganisms possess pigments like chlorophyll, bacteriochlorophyll, cytochromes, and flavins, which can absorb light energy, become excited or activated, and act as photosensitizers. The excited photosensitizer (P) transfers its energy to O_2 generating singlet oxygen ($^1O_2$).

$$P \xrightarrow{\text{light}} P \text{ (activated)}$$

$$P \text{ (activated)} + O_2 \longrightarrow P + {}^1O_2$$

Singlet oxygen is a very reactive, powerful oxidizing agent that will quickly destroy a cell. It is probably the major agent employed by phagocytes to destroy engulfed bacteria. Many microorganisms that are airborne or live on exposed surfaces use carotenoid pigments for protection against photooxidation. Carotenoids effectively quench singlet oxygen—that is, they absorb energy from singlet oxygen and convert it back into the unexcited ground state. Both photosynthetic and nonphotosynthetic microorganisms employ pigments in this way.

5 Microbial Growths in Natural Environments

The previous section surveyed the effects on microbial growth of individual environmental factors such as water availability, pH, and temperature. Although microbial ecology will be introduced in more detail at a later point, we will now briefly consider the effect of the environment as a whole on microbial growth.

Growth Limitation by Environmental Factors

The microbial environment is complex and constantly changing. Characteristically microorganisms in a particular location are exposed to many overlapping gradients of nutrients and various other environmental factors. This is particularly true of microorganisms growing in biofilms. Microorganisms will grow in "microenvironments" until an environmental or nutritional factor limits growth. Liebig's law of the minimum states that the total biomass of an organism will be determined by the nutrient present in the lowest concentration relative to the organism's requirements.

This law applies in both the laboratory (figure 2) and in terrestrial and aquatic environments. An increase in a limiting essential nutrient such as phosphate will result in an increase in the microbial population until some other nutrient becomes limiting. If a specific nutrient is limiting, changes in other nutrients will have no effect. The situation may be even more complex than this. Multiple limiting factors can influence a population over time. Furthermore, as we have seen, factors such as temperature, pH, light, and salinity influence microbial populations and limit growth.

Shelford's law of tolerance states that there are limits to environmental factors below and above which a microorganism cannot survive and grow, regardless of the nutrient supply. This can readily be seen for temperature in figure 5. Each microorganism has a specific temperature range in which it can grow. The same rule applies to other factors such as pH, oxygen level, and hydrostatic pressure in the marine environment. The growth of a microorganism depends on both the nutrient supply and its tolerance of the environmental conditions.

Most microorganisms are confronted with deficiencies that limit their activities except when excess nutrients allow unlimited growth. Such rapid growth will quickly deplete nutrients and possibly result in the release of toxic waste products, which will limit further growth. In response to low nutrient levels (oligotrophic environments) and intense competition, many microorganisms become more competitive in nutrient capture and exploitation of available resources. Often the organism's morphology will change in order to increase its surface area and ability to absorb nutrients. This can involve conversion of rod-shaped procaryotes to "mini" and "ultramicro" cells or changes in the morphology of prosthecate prokaryotes, in response to starvation. Nutrient deprivation induces many other changes as discussed previously. For example, microorganisms can undergo a step-by-step shutdown of metabolism except for housekeeping maintenance genes.

Many factors can alter nutrient levels in oligotrophic environments. Microorganisms may sequester critical limiting nutrients, such as iron, making them less available to competitors. The atmosphere can contribute essential nutrients and support microbial growth. This is seen in the laboratory as well as natural environments. Airborne organic substances have been found to stimulate microbial growth in dilute media, and enrichment of growth media by airborne organic matter can allow significant populations of microorganisms to develop. Even distilled water, which already contains traces of organic matter, can absorb one carbon compounds from the atmosphere and grow microorganisms. The presence of such airborne nutrients and microbial growth, if not detected, can affect experiments in biochemistry and molecular biology, as well as studies of microorganisms growing in oligotrophic environments.

Natural substances also can directly inhibit microbial growth and reproduction in low-nutrient environments. These agents include phenolics, tannins, ammonia, ethylene, and volatile sulfur compounds. This may be a means by which microorganisms avoid expending limited energy reserves until an adequate supply of nutrients becomes available. Such chemicals are also important in plant pathology and may aid in controlling soil-borne microbial diseases.

Quorum Sensing and Microbial Populations

For decades microbiologists tended to think of bacterial populations as collections of individuals growing and behaving independently. More recently it has become clear that many bacteria can communicate with one another and behave cooperatively. A major way in which this cooperation is accomplished is by a process known as quorum sensing or autoinduction. This is a phenomenon in which bacteria monitor their own population density through sensing the levels of signal molecules, sometimes called autoinducers because they can stimulate the cell that releases them. The concentration of these signal molecules increases along with the bacterial population until it rises to a specific threshold and signals the bacteria that the population density has reached a critical level or quorum. The bacteria then begin expressing sets of quorum-dependent genes.

Quorum sensing has been found among both gram-negative and gram-positive bacteria. Quorum sensing makes good practical sense. Take the production and release of extracellular enzymes as an example. If such enzymes were released by only a few bacteria, they would diffuse away and be rendered ineffective because of dilution. With control by quorum sensing, the bacteria reach a high population density before they release enzymes, and as a consequence enzyme levels are concentrated enough to have significant effects. This is an advantage within a host's body as well as in the soil or an aquatic habitat. If a pathogen can reach high levels at a particular site before producing virulence factors and escaping into surrounding host tissues, it has a much better chance of counteracting host defenses and successfully spreading throughout the host's body. This explains another pattern in quorum sensing. It seems to be very important in many bacteria that establish symbiotic or parasitic relationships with hosts.

Quorum sensing was first discovered in gram-negative bacteria and is best understood in these microorganisms. The most common signals in gram-negative bacteria are acyl homoserine lactones (HSLs). These are small molecules composed of a 4- to 14-carbon acyl chain attached by an amide bond to homoserine lactone (figure 8a). The acyl chain may have a keto group or hydroxyl group on its third carbon. Acyl HSLs diffuse into the target cell (figure 8b). Once they reach a sufficiently high level, acyl HSLs bind to special receptor proteins and trigger a conformational change. Usually the activated complexes act as inducers—that is, they bind to target sites on the DNA and stimulate transcription of quorum-sensitive genes. The gene needed to synthesize acyl HSL is also produced frequently, thus amplifying the effect by the production and release of more autoinducer molecules.

Many different processes are sensitive to acyl HSL signals and quorum sensing in gram-negative bacteria. Some well-studied examples are (1) bioluminescence production by Vibrio fischeri, (2) Pseudomonas aeruginosa synthesis and release of virulence factors, (3) conjugal transfer of genetic material by Agrobacterium tumefaciens, and (4) antibiotic production by Erwinia carotovora and Pseudomonas aureofaciens. Gram-positive bacteria also regulate activities by quorum sensing, often using an oligopeptide signal. Good examples are mating in Enterococcus faecalis, competence induction in Streptococcus pneumoniae, stimulation of sporulation by Bacillus subtilis, and production of many toxins and other virulence factors by Staphylococcus aureus. Quorum sensing even stimulates the development of aerial mycelia and the production of streptomycin by *Streptomyces griseus.* In this case, the signal seems to be γ-butyrolactone rather than an oligopeptide.

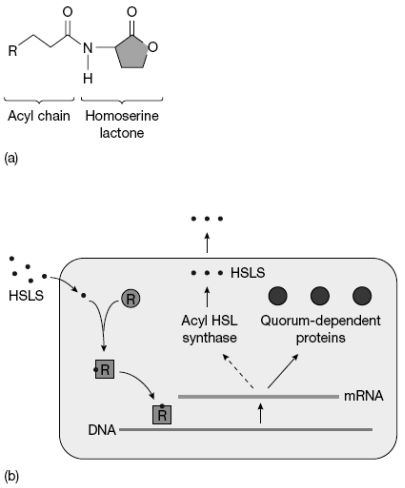


Figure 8 Quorum Sensing in Gram-Negative Bacteria. (**a**) A generalized structure for acyl homoserine lactone, the best-known quorum sensing signal or autoinducer. (**b**) A schematic diagram giving an overview of the way in which quorum sensing functions in many gram-negative bacteria. The receptor protein that acts as an inducer is labeled R. The dashed lines indicate that acyl HSL synthase is not always made in response to the autoinducer. See text for more details.

An interesting and important function of quorum sensing is to promote the formation of mature biofilms by the pathogen Pseudomonas aeruginosa, and it may play a role in cystic fibrosis. Biofilm formation makes sense for the pathogen because biofilms protect against antibiotics and detergents. Quorum sensing should be very effective within biofilms because there will be less dilution and acyl HSL levels will increase rapidly. Under such circumstances, two different bacteria might stimulate each other by releasing similar signals; this appears to be the case in biofilms containing the pathogens P. aeruginosa and Burkholderia cepacia. Quorum sensing is an example of what might be called multicellular behavior in that many individual cells communicate and coordinate their activities to act as a unit. Other examples of such complex behavior are pattern formation in colonies and fruiting body formation in the myxobacteria.

UNIT-3

FUNDAMENTAL PROCESSES

INSTITUTE FOR ADVANCED STUDIES

B-7 SARASWATI NAGAR, JODHPUR

e-mail:ifasnet@gmail.com

http://csirnetlifesciences.tripod.com

BLANK PAGE

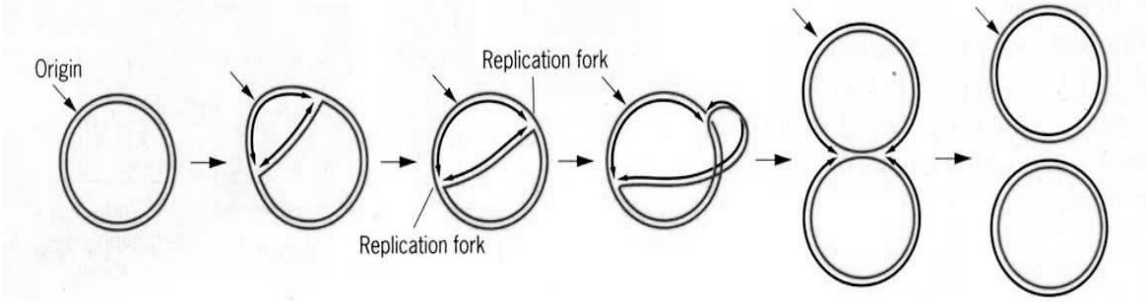
A1. DNA replication: Unit of replication, enzymes involved, replication origin and replication fork, fidelity of replication

DNA Replication in Prokaryotes:

DNA replication is carried out be enzymes called DNA polymerases using deoxyribonucleoside triphosphates (dNTPs) as precursors. The process starts with the separation of the two strands of the DNA helix to expose the bases. New strands of DNA are synthesized using the old strands as templates. Complementary base-pairing between the bases on the incoming nucleotides and the template DNA strand (A:T; G:C) dictates which nucleotide will be inserted into the growing DNA chain. This is called **semiconservative replication** as, after the whole DNA molecule has been copied, each new DNA helix consists of one original DNA strand and one newly synthesized DNA strand. The role of the DNA polymerase is to check that the hydrogen bonding is correct and then to form a phosphodiester bond between two adjacent nucleotides. As the basic mechanism for DNA replication is the same for all organisms, this will be described first in relation to bacterial DNA replication. Differences found in eukaryotes will be discussed subsequently.

Origin of Replication:

The first stage of replication involves the identification of the origin of replication. There is only one on the chromosome of bacteria called ***oriC***. A complex of proteins binds to this site, opens up the helix and initiates replication. DNA synthesis occurs in both directions (bidirectional) and on both strands, creating a replication bubble which appears as a θ (theta) structure in electron micrographs. The two sites at which DNA synthesis occurs are called the replication forks. As replication proceeds the replication forks move round the molecule opening up the DNA strands, synthesizing two new complementary strands and rewinding up the DNA behind until the two replication forks meet at a termination site. The two completed circles of DNA are still linked and have to be separated by an enzyme **topoisomerase II** which makes a transient break in both strands of one of the molecules, allowing them to disconnect.



Other enzymes/proteins associated with DNA replication are

- **DNA helicase** which unwinds the DNA helix to allow access to the bases.
- **Single-stranded binding (SSB) proteins** that bind to the DNA strands, preventing them from base-pairing before they have been copied.
- There is a problem in that as the helix is opened up the DNA ahead of the replication fork would become overwound because the DNA ends are fixed the chromosome is circular) and therefore cannot rotate relative to each other. This is solved by **topoisomerase I** which makes a break in one of the DNA strands which allows the DNA to rotate round the intact strand thus releasing the torsion.

DNA Polymerase: enzyme responsible for normal DNA synthesis in bacteria is DNA polymerase III. DNA polymerase I also has a role in DNA replication and is the enzyme that is involved in DNA repair. The role of DNA polymerase II is unknown as yet. These enzymes function by adding a new nucleotide to the 3' -OH end of a growing polynucleotide chain with the release of pyrophosphate synthesizing a new strand of DNA in a 5'3' direction.

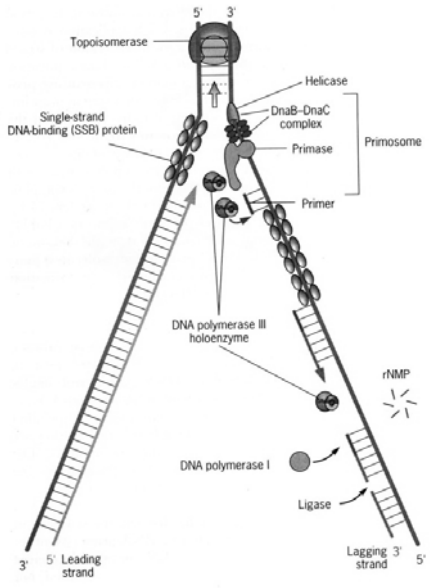


Table : Some *E. coli* replication-deficient mutants and the functions of each locus

Locus	Phenotype	Function
dnaA	Slow-stop	Initiator protein
dnaB	Quick-stop	Helicase
dnaC	Quick- and slow-stop alleles	Forms complex with DnaB helicase and assists loading at origin
dnaE (*polC*)	Quick-stop	α subunit of pol III holoenzyme
dnaG	Quick-stop	Primase
dnaN	Quick stop	β unit of pol III holoenzyme
dnaQ	Quick stop and mutator alleles	ε subunit of pol III holoenzyme
dnaT	Slow-stop	Primosome component
dnaX	Quick stop	γ and τ subunits of pol III holoenzyme
gyrA	Quick- and slow-stop alleles	DNA gyrase subunit α
gyrB	Quick- and slow-stop alleles	DNA gyrase subunit β
lig	Repair deficient	DNA ligase
ori	Lethal (*not conditional*)	Origin of replication
polA	Repair deficient	DNA polymerase I
rnhA	Stable replication	RNase H
ssb	Quick-stop	Single-strand binding protein
ter	None	Terminus of replication
tus	None	Termination protein

DNA Fidelity/Proof reading:

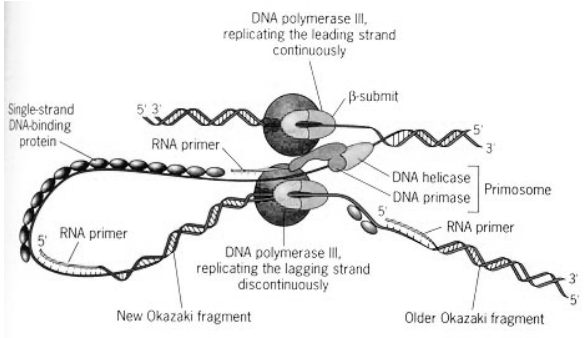
DNA polymerases also have exonuclease activity which allows the sequential removal of nucleotides in both the 5'→3' and 3'→5' directions. The ability to remove incorrectly inserted nucleotides confers proofreading properties on DNA polymerase III. Immediately after the enzyme has inserted a nucleotide into the ring chain the enzyme can detect if the base-pairing is correct or if a mismatch has occurred. If there is a mismatch, the 3'→5' exonuclease activity removes the nucleotide and the polymerase can then replace it with the correct one. This proof-reading ability ensures that the error rate (mutation rate) in synthesis is therefore less than **10^{-8}-10^{-10} per base pair**.

There are several features of DNA replication that are due to the nature of all polymerase enzymes.

1. DNA polymerases cannot start from scratch - they must have a 3'-OH group to add a nucleotide on to. These enzymes therefore require some sort of primer to start DNA synthesis.
2. DNA can only be synthesized in one direction. 5'→3'. As the two strands DNA are antiparallel and are both copied at the same time at the replication fork only one strand can be synthesized continuously. The other strand must be synthesized in short fragments.

Primers: The problem of how to start DNA synthesis is solved by using a specialist RNA polymerase called a **primase** to lay down a short stretch of ribonucleotides (primer) complementary to the DNA template at the start point for DNA synthesis. RNA polymerases do not require a 3'-OH group to start RNA synthesis so can start from scratch. The short piece of RNA provides the necessary 3' -OH group to which DNA polymerase ill can join a deoxyribonucleotide and start to synthesize the DNA strand. Synthesis continues until the DNA polymerase III reaches a primer. DNA polymerase III is released and DNA polymerase I binds in its place. DNA polymerase I has a 3'→5' exonuclease activity which confers the ability to remove the ribonucleotides one by one while its polymerase activity allows it to replace these with deoxyribonucleotides. The final gap in the deoxyribonucleotide Chain is sealed by a specialist enzyme, **DNA** ligase, which can join adjacent 5'-phosphate and 3'-OH groups.

As DNA can only be synthesized in the 5'→3' direction, only one strand can be synthesized continuously. This is called the **leading strand**. The other strand, called the lagging strand, is synthesized as short pieces, approximately 1000 bases long called Okazaki fragments. On this strand there is, therefore, constant re-initiation of DNA synthesis by a complex of proteins called the primosome, as described in the section above, with the consequent joining of the fragments by the removal of the RNA primers by DNA polymerase I and sealing of the gaps by DNA ligase.



Rolling Circle replication:

An alternative mechanism for DNA replication is used by some bacteriophage and viruses and in the transfer of plasmid DNA between cells by conjugation called rolling-circle replication, one strand of the DNA is nicked to provide 3′-OH group. Helicase and SSB proteins create a replication fork and nicked strand is displaced. dNTPs are added to the 3′-OH group to create the leading strand and the displaced strand acts as a template for lagging strand synthesis.

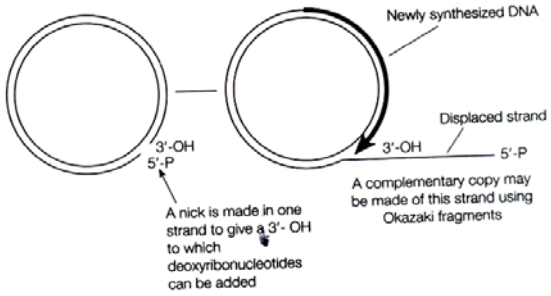


Fig. Diagram showing the mechanism of circle replication.

Eukaryotic Replication:

The main differences between prokaryotic and eukaryotic replication are mainly as a result of the larger size of the genomes and the fact that they are linear. Some of the main ones are listed below:

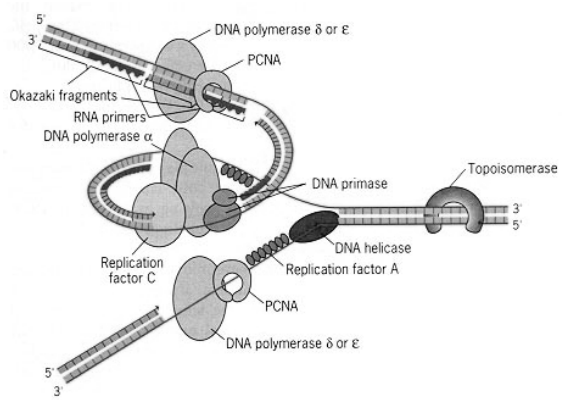
- There are many origins of replication on one chromosome, approximately one every 3-300 kb of DNA.
- Leading and lagging strands are made by different DNA polymerases. The Okazaki fragments are shorter.
- Because of the need to reinitiate DNA synthesis constantly on the lagging strands there are short segments at the ends (telomeres) which do not get copied. The chromosomes would therefore gradually shorten with time. This problem is remedied by a specialist RNA-containing enzyme, a **telomerase**, which extends the ends of chromosomes by adding numerous hexanucleotide repeat sequences. These then act as templates for DNA replication.

Because of the complexities of chromatin structure, eukaryotic replication forks move at around 50 bp per sec. At this rate, it would take about 30 days to copy a typical mammalian chromosomal DNA molecule of 10^8 bp using two replication forks, hence the need for multiple replicons - typically 50000-100000 in a mammalian cell. To study all these simultaneously would be a daunting prospect. Fortunately, the yeast *Saccharomyces cerevisiae* has a much smaller genome (1.4×10^7 bp in 16 chromosomes) and only 400 replicons. The complete genome sequence is known. Simpler still are viruses such as simian virus 40 (SV40). SV40 DNA is a 5 kb double-stranded circle which forms nucleosomes when it enters the cell nucleus. It provides an excellent model for a mammalian replication fork.

Another useful system is a cell-free extract prepared from the eggs of the African clawed frog, *Xenopus laevis.* Because of its high concentration of replication proteins (sufficient for several cell doublings *in vivo),* this extract can support the extensive replication of added DNA or even whole nuclei.

Origin and Initiation:

In eukaryotes, clusters (tandem arrays) of about 20-50 replicons initiate simultaneously at defined times throughout S-phase. Those which replicate early in S-phase comprise predominantly euchromatin (which includes transcriptionally active DNA) while those activated late in S-phase are mainly within heterochromatin. Centromeric and telomeric DNA replicates last. This pattern reflects the accessibility of the different chromatin structures to initiation factors. Individual yeast replication origins have been cloned into prokaryotic plasmids. Since they allow these plasmids to replicate in yeast (a eukaryote) they are termed **autonomously replicating sequences (ARSs)**. The minimum length of DNA that will support replication is only 11 bp and has the consensus sequence [A/T]TTTAT[A/G]TTT[A/T], though additional copies of this sequence are required for optimal efficiency. This sequence is bound by the **origin recognition complex** (ORC) which, when activated by CDKs, permits opening of the DNA for copying. Defined origin sequences have not been isolated from mammalian cells, and it is believed that initiation of each replicon may occur at random within an initiation zone which may be several kilobases in length and which may be part of the dispersed repetitive DNA fraction. In contrast to prokaryotes, eukaryotic replicons can only initiate once per cell cycle. A protein (licensing factor) which is absolutely required for initiation and inactivated after use can only gain access to the nucleus when the nuclear envelope dissolves at mitosis, thus preventing premature re-initiation.



Replication Fork:

Before copying, the DNA must be unwound from the nucleosomes at the replication forks. This slows the movement of the forks to about 50 bp per sec. After the fork has passed, new nucleosomes are assembled from

a mixture of old and newly synthesized histones. As in prokaryotes, one or more DNA helicases and a single-stranded binding protein, **replication protein A (RP-A)**, are required to separate the strands. Three different DNA polymerases are involved in elongation. The leading strands and each lagging strand fragment are initiated with RNA by a primase activity that is an integral part of **DNA polymerase α**. This polymerase continues elongation with DNA but is quickly replaced by **DNA polymerase-δ** (delta) on the leading strand and probably also on the lagging strand. The role of **DNA polymerase-ε** (epsilon) is less clear, but it may simply complete the lagging .strand fragments after primer removal. Both these enzymes have associated proofreading activity. The ability to synthesize long DNA is conferred on DNA polymerase δ by **proliferating cell nuclear antigen (PCNA)**, the functional equivalent of the β subunit of E. *coli* DNA polymerase III holoenzyme. The small size of eukaryotic lagging strand fragments (e.g. 135 bp in SV40) appears to reflect the amount of DNA unwound from each nucleosome as the fork progresses. In addition to doubling the DNA, the histone content of the cell is also doubled during S-phase.

Table : Properties and proposed functions of eukaryotic DNA polymerases

Mammalian name	α	β	γ	δ	ε
Yeast name	pol1	pol4	polM	pol3	pol2
Yeast gene	*POL1*	*POL4*	*MIP1*	*POL3*	*POL2*
Location	Nuclear	Nuclear	Mitochondrial	Nuclear	Nuclear
Number of subunits	4	1	2	2	>1
5′→3′ polymerase	√	√	√	√	√
3′→5′ exonuclease	x	x	√	√	√
Primase	√	x	x	x	x
Associated factors	None	None	None	PCNA	None
Processivity	Moderate	Low	High	High with PCNA	High
Function	Lagging-strand priming	Repair polymerase	Organelle polymerase	Principle replicative polymerase	Unknown

Nuclear Matrix:

The nuclear matrix is a scaffold of insoluble protein fibers which acts as an organizational framework for nuclear processes, including DNA replication. Huge replication factories containing all the enzymes and DNA associated with the replication forks of all replicons within a cluster are immobilized on the matrix, and ~he DNA moves through these sites as it replicates. These factories can be visualized in the microscope by pulse-labeling the replicating DNA with the thymidine analog, **bromodeoxyuridine (BUdR)**, and visualizing the labeled DNA by immunofluorescence using an antibody that recognizes BUdR..

Telomeric Replication:

The ends of linear chromosomes cannot be fully replicated by semi-discontinuous replication as there is no DNA to elongate to replace the RNA removed from the 5'-end of the lagging strand. Thus, genetic information could be lost from the DNA. To overcome this, the ends of eukaryotic chromosomes consist of hundreds of copies of a simple, non informational repeat sequence (e.g. TTAGGG in humans) with the 3'-end overhanging the 5'-end.

3'-AATCCCAATCCC-5'

5'-TTAGGGTTAGGG---(TTAGGG)$_n$----TTAGGG-3'

The enzyme telomerase contains a short RNA molecule, part of whose sequence is complementary to this repeat. This RNA acts as a template for the addition of these repeats to the 3'-overhang by repeated cycles of elongation (polymerization) and translocation. The complementary strand is then synthesized by normal lagging strand synthesis, leaving a 3'-overhang. Interestingly, telomerase activity is repressed in the somatic cells of multicellular organisms, resulting in a gradual shortening of the chromosomes with each cell generation. As this shortening reaches informational DNA, the cells senesce and die. This phenomenon may be important in cell aging. Furthermore, the unlimited proliferative capacity of many cancer cells is associated with reactivation of telomerase activity.

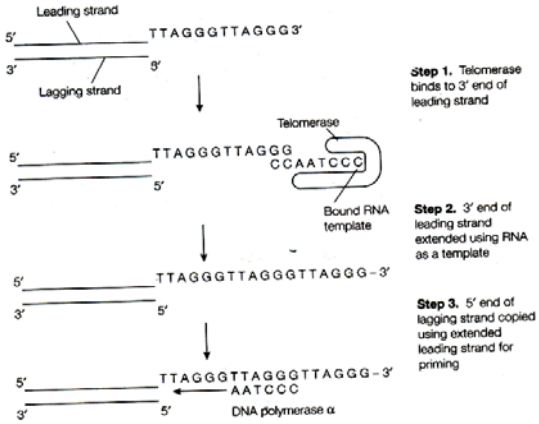


Fig. Replication of DNA at telomeres.

A2. EXTRA-CHROMOSOMAL REPLICONS

Plasmid replication and maintenance

Plasmid copy number: Plasmid replication and maintenance is a totally intrinsic property applicable to all plasmids, allowing them to be classified according to replication and partition strategy and how their copy number is regulated.

Plasmid replication is autonomous, but may be coupled to the replication of the host genome and may be influenced by the host and its environment. Generally, plasmids require use of the host encoded replication machinery (e.g. DNA polymerase, RNA polymerase, DNA primase), but encode other factors required for the initiation of replication and its regulation. Initiation occurs at a specific origin and is the rate-limiting step in plasmid replication. The frequency of successful initiation establishes the characteristic average number of plasmid molecules per cell, the copy number.

In bacteria, larger plasmids usually have a copy number of 1-3: their replication is coupled to that of the chromosome. These are termed single-copy plasmids and they are usually conjugative. Smaller plasmids have higher copy numbers, typically 10-30, and are termed multi copy plasmids.

Table 1: Definition of some terms used in plasmid biology

Term	Definition
Basic replicon	The minimal region of a plasmid able to replicate in the same manner as the full-sized plasmid — typically contains the origin of replication and **plasmid maintenance sequences**, elements that regulate replication and plasmid *maintenance*
Conjugative and nonconjugative plasmids	Plasmids which can, or cannot, promote their own transfer by *conjugation*
Copy number	The average number of plasmids per cell, usually measured by direct quantification of plasmid DNA or an encoded gene product
***cop* mutant**	A plasmid carrying a mutation which affects copy number
Cryptic plasmid	A plasmid with no apparent phenotype
Curing	Spontaneous or induced plasmid loss. Spontaneous curing occurs at low frequency, but can be induced by intercalating agents and some antibiotics
Dislodgment	The rare displacement of a resident plasmid by a second, *compatible* plasmid, a phenomenon often interpreted as *incompatibility* (q.v.), but which may reflect, e.g., the activity of a restriction endonuclease encoded by the second plasmid
Episome	A plasmid or virus capable of both extrachromosomal replication and integration into the host genome, e.g. the F-plasmid (*see* Gene Transfer in Bacteria)
Homoplasmid, heteroplasmid	Describing cells containing one type of plasmid, or two distinct types of plasmid (q.v. *plasmid segregation*)
Incompatibility	The inability of two different types of plasmid to coexist in the same cell for more than a few generations in the absence of selection for both plasmids, reflecting common replication or partition mechanisms
Invertron	A linear plasmid in eukaryotes with long, perfect, inverted terminal repeats
Killer system	A maintenance system which ensures that cured daughter cells are destroyed
Maintenance system	A system which ensures that plasmids are maintained in a population of dividing cells
Miniplasmid	The basic replicon of a large plasmid such as the F-plasmid
One-way incompatibility	A phenomenon where the introduction of plasmid type 1 into a population of cells where plasmid type 2 is established results in incompatibility, but introduction of plasmid type 2 into a population where plasmid type 1 is established does not. This reflects similar but not identical replication mechanisms in the two plasmids
Partition, partition system	Partition is the distribution of plasmids into daughter cells during cell division. Partition systems are maintenance systems which ensure equal partition
Plasmid	An autonomous, extrachromosomal replicon which is nonessential under normal growth conditions and not part of the cellular genome
Plasmid origin	The plasmid locus where DNA replication begins
Plasmid segregation	The separation of different types of plasmid into separate daughter cells at cell division (due to incompatibility)
Promiscuous plasmid (broad host-range plasmid)	A bacterial plasmid with a broad host range (usually including both gram-positive and gram-negative bacteria)
Prime plasmid	A plasmid episome which has excised aberrantly and carries part of the host chromosome. **Type I prime plasmids** have exchanged plasmid genes for host genes, **type II prime plasmids** carry all plasmid genes plus extra host genes (*F′ plasmid*)
Relaxed plasmid	A plasmid whose replication does not require continued protein synthesis and whose copy number increases if protein synthesis is blocked, due to the removal of a negative regulator protein
Stringent plasmid	A plasmid whose replication requires continued protein synthesis and whose copy number thus falls if protein synthesis is inhibited

Replication occurs randomly and is self-regulated. The plasmids are usually nonconjugative. The replication control mechanisms which maintain the correct copy number do so by measuring the relative concentration of origins in the cell. If a mutation disrupts the function of a key regulator, the copy number can change substantially. Severely repressed replication can lead to plasmid loss, and unregulated replication (runaway replication) can increase the copy number 10-fold.

Runaway replication is capped only when another component, e.g. an enzyme or substrate for DNA replication, becomes limiting. A plasmid carrying such a mutation is known as a ***cop*** **mutant.**

The behavior of eukaryotic plasmids has been studied in the yeast ***S. cerevisiae.*** Plasmids with the 2μ origin are maintained at high copy number (<100), although replication occurs only in the ***S-phase*** and is coupled to genome replication. Conversely, plasmids which resemble chromosomes (i.e. those possessing ARS origins and a centromere) are maintained at a copy number of 1-2. The centromere is dominant to plasmid origins, so that artificial plasmids containing centromeres, ARSs and 2μ origins behave like chromosomes (***yeast cloning vectors).***

Plasmid maintenance: As well as controlling self-transfer and replication, plasmids also encode maintenance functions, ensuring both daughter cells inherit plasmids following cell division. The distribution of plasmids to the daughters of a dividing cell is partition. Without maintenance, spontaneous ***curing*** (Table 1) can occur during partition. Cells without plasmids are more competitive than plasmid containing cells under normal growth conditions (i.e. in the absence of selection for the phenotype conferred by the plasmid) because they do not divert their resources to plasmid functions- therefore, cured cells rapidly increase in the population at the expense of the plasmid containing cells. Maintenance functions have been well characterized in bacteria and fall into several groups ***(Table 2).*** The existence of maintenance functions places plasmids, along with viruses and transposable elements, within the category of ***selfish DNA***.

Table 2 : Plasmid maintenance functions

Maintenance function	Description
Partition (*par*) system	A system found in low-copy number plasmids that accurately and equally distributes plasmid copies to either side of the cell prior to division. Partition systems require both *trans*-acting factors and a *cis*-acting element located on the plasmid, and may involve attachment of the plasmid to the cell membrane. The bacterial chromosome uses a similar partition mechanism . High copy number plasmids lack specific partition mechanisms and rely on the high probability that each daughter will receive at least one copy
Killer (*kil*) systems (addiction systems)	A system in which the plasmid encodes a stable **killer protein** (a protein which is lethal to the host cell) and an unstable 'antidote' molecule acting as an antagonist of the killer protein itself or an inhibitor of its synthesis. In cured cells, the killer protein outlasts the antidote and the cell is killed
Cell division delay	A system which delays cell division at low plasmid copy number. The F-plasmid is able to delay cell division by inducing the *SOS response* .
Recombination systems	The plasmid encodes a *site-specific recombination system* which counteracts homologous recombination events leading to multimerization and ensures that monomers are available for partition. A similar function, encoded by the *xerC* and *xerD* genes of *E. coli*, ensures chromosome monomerization. The yeast 2μ plasmid also encodes a site-specific recombination system which increases its copy number

Note that partition systems, killer systems and cell division delay systems ensure that plasmid segregation is better than random by preventing the birth of cured cells. Conversely, recombination systems do not prevent the birth of cured cells, and they can only achieve at best random plasmid segregation.

Plasmid incompatibility: Incompatibility is the inability of two plasmids to coexist in the same host strain unless conditions are imposed which select for the phenotype conferred by both plasmids.

Incompatible plasmids rapidly segregate in a growing unselected population to yield two homoplasmid strains. Incompatibility occurs between plasmids with similar strategies for the control of replication and/or partition. The control of plasmid copy number is predominantly by negative regulation; thus two distinct single-copy plasmids with the same regulatory mechanism in the same cell will repress each other's replication until cell division, when they become segregated and repression is lifted. Mixed multicopy plasmids also demonstrate mutual repression until cell division. In the daughter cells, repression is lifted, and because multicopy plasmid replication is random, each type of plasmid has an equal chance to undergo replication. The type which succeeds will then achieve a higher copy number and will be at an advantage at the next round of division, eventually leading to the generation of homoplasmid cells. Multicopy plasmids thus take longer to

segregate than single-copy plasmids. Single-copy plasmids with the same partition mechanism segregate due to competition for the same 'partition site', which is presumably represented only twice in the cell.

Plasmid incompatibility provides a useful system of classification according to replication and partition strategy. Mutually incompatible plasmids (i.e. those with similar strategies) are placed into an incompatibility group (Inc group), of which there are approximately 30 for *E. coli* and related enterobacteria.

Mechanisms of plasmid DNA replication:

Most of the typical closed circular double-stranded DNA plasmids of bacteria replicate similarly to the chromosome, with initiation characterized by the binding of specific initiation proteins called Rep proteins to repetitive elements at the plasmid origin. For many plasmids (e.g. F, R1), Rep facilitates unwinding of the origin, allowing the loading of helicase and the establishment of either a single replication fork, or two replication forks. In other plasmids (e.g. pT181 of *Staphylococcus),* the Rep protein is a *nickase* which initiates *rolling circle replication*. ColE1-related plasmids do not require Rep proteins because the host RNA polymerase is used to transcribe through the origin to generate a primer for the leading strand; such plasmids are thus under *relaxed control*.

Linear bacterial plasmids are similar to linear virus genomes, e.g. they possess covalently sealed ends (e.g. ***Borellia*** plasmids) or terminal proteins and inverted repeats (e.g. *Streptomyces* plasmids) and may replicate using the same strategies employed by viruses Linear plasmids of eukaryotic organelles often encode their own polymerases, which are presumably utilized for autonomous replication. The discussion below relates to strategies for the regulation of replication unique to plasmids.

Multiple origins and iterons:

Many larger plasmids have multiple origins (often associated with different control mechanisms), making it difficult to assign such plasmids to single incompatibility groups. In some plasmids (e.g. the F plasmid), one particular origin is favored, and the role of the extra origins is uncertain. In others (e.g. R6K) the alternative origins are used with equal frequency.

Many recombinant plasmid vectors have been deliberately designed to incorporate both prokaryotic and eukaryotic origins, allowing them to be maintained in both types of cells (*shuttle vector).*

Plasmid origins are often characterized by essential repetitive sequences termed iterons. These are common motifs in plasmids (e.g. plasmid RK2) and in some phages which can replicate like plasmids (e.g. bacteriophages P1 and λ*),* as well as the bacterial chromosome itself (*origin of replication).*

Plasmid iterons are binding sites for the Rep proteins, which may act alone or may associate with host initiator proteins (e.g. DnaA in *E. coli).* In plasmid RK6, for instance, regardless of which of the three origins is used, a sequence of seven direct tandem repeats located within the γ-origin is essential for replication.

Control of plasmid replication** by **antisense RNA:

Plasmids with a ColEl type origin of replication (including the vast majority of plasmid cloning vectors) initiate leading-strand DNA synthesis from a single RNA primer generated by the host RNA polymerase (*primer, priming strategy).* The origin **is** actually transcribed on both strands to yield two transcripts, RNA I and RNA II. RNA II is the primer, whereas RNA I, which is complementary to part of RNA II, acts as a repressor by sequestering RNA II into an inactive duplex. DNA synthesis begins at the origin, but transcription of RNA II begins 555 bp upstream of the origin and continues through it and beyond. The transcript must therefore be processed by cleavage at the origin to yield a functional primer. This processing is carried out by RNaseH (**nucleases)** and is facilitated by the three-stem loop secondary structure adopted by RNA II.

Interaction between RNA I and RNA I1 culminates in their full ***hybridization***, which disrupts the secondary structure of RNA II and prevents it hybridizing to the DNA. This in turn prevents the cleavage reaction and results in a replication block. ***In vitro,*** RNA I is able to anneal to RNA II only when the latter is between 100 and 360 nt in length. This means that RNA I must hybridize to the nascent RNA II transcript during the early phase of its synthesis to repress replication. A plasmid encoded protein called Rom enhances the rate of RNA I: RNA II hybrid formation once RNA II is greater than 200 nucleotides in length and is thus a key regulator of ColE1 replication.

The replication of the R1 plasmid is also regulated by antisense RNA. The plasmid-encoded initiator protein, RepA, is essential for the initiation of replication, but transcription of an adjacent gene generates an antisense RNA (copA) which binds to the leader region of the **repA** mRNA (designated copT) to prevent its translation (***antisense RNA, translational control).*** The copA and *copT* RNAs both form secondary structures, and mutational analysis has shown that the interaction between them may involve these secondary structures rather than full duplex formation (a so-called kissing complex).

The prevalence of antisense RNA regulators in plasmids suggests that RNA offers a unique strategy for replication control for which protein is insufficient. In many diverse plasmids the same mechanism occurs: a short antisense inhibitory factor anneals to the 5' end of an essential functional RNA

and blocks its activity. A possible explanation is that the maintenance of plasmid copy number requires an unstable regulator so that the plasmid population can respond rapidly to deviations from the standard copy number. RNA is a suitable candidate for such an unstable regulator (***antisense RNA).***

Relaxed and stringent control:

Plasmid replication may be under relaxed or stringent control. If cellular protein synthesis is inhibited, for example by chloramphenicol, some plasmids continue to replicate and their copy number increases above normal vegetative levels. These are relaxed plasmids, and are typically small multicopy plasmids. Other plasmids cease to replicate along with the host chromosome when ***de novo*** protein synthesis is inhibited. These are described as stringent plasmids, and are typically large, single-copy, conjugative plasmids. Most general purpose plasmid cloning vectors are under relaxed control.

The ability of relaxed plasmids to continue to replicate in the absence of host protein synthesis reflects the stability of host proteins required for the initiation of replication, and the instability of plasmid-encoded regulators. For instance, plasmid ColE1 is under relaxed control because the initiation factor Rom has a negative role - it prevents initiation by facilitating the pairing of an antisense RNA with the leading stand primer. In the absence of Rom, plasmid replication is de-repressed and copy number increases (very high number cloning vectors have no functional **rom** gene). Conversely, stringent plasmids such as F require a positively acting Rep protein for initiation. In the absence of Rep, plasmid replication is blocked regardless of the availability of host DNA replication proteins.

A3. DNA DAMAGE AND THEIR ROLE IN CARCINOGENESIS

The DNA sequence can be changed as the result of copying errors introduced by DNA polymerases during replication and by environmental agents such as mutagenic chemicals and certain types of radiation. If DNA sequence changes, whatever their cause, are left uncorrected, both growing and non-growing somatic cells might accumulate so many mutations that they could no longer function. In addition, the DNA in germ cells might incur too many mutations for viable offspring to be formed. Thus the correction of DNA sequence errors in all types of cells is important for survival.

The relevance of DNA damage and repair to the generation of cancer (carcinogenesis) became evident when it was recognized that all agents that cause cancer (carcinogens) also cause a change in the DNA sequence and thus are mutagens. All the effects of carcinogenic chemicals on tumor production can be accounted for by the DNA damage that they cause and by the errors introduced into DNA during the cells' efforts to repair this damage. Likewise, ultraviolet (UV) radiation and ionizing radiation (x-rays and atomic particles) not only modify DNA, but also can cause cancer in animals and can transform normal cells in culture into rapidly proliferating, cancer-type cells. The ability of ionizing radiation to cause human cancer, especially leukemia, was dramatically shown by the increased rates of leukemia among survivors of the atomic bombs dropped in World War II, and more recently by the increase in melanoma (skin cancer) in individuals exposed to too much sunlight.

Lesion: A lesion is an alteration to the normal chemical or physical structure of the DNA. Some of the nitrogen and carbon atoms in the heterocyclic ring systems of the bases and some of the exocyclic functional groups (i.e. the keto and amino groups of the bases) are chemically quite reactive. Many exogenous agents, such as chemicals and radiation, can cause changes to these positions. The altered chemistry of the bases may lead to loss of base pairing or altered base pairing (e.g. an altered A may base-pair with C instead of T). If such a lesion was allowed to remain in the DNA, a mutation could become fixed in the DNA by direct or indirect mutagenesis. Alternatively, the chemical change may produce a physical distortion in the DNA which blocks replication and/or transcription, causing cell death. Thus, DNA lesions may be mutagenic and/or lethal. Some lesions are spontaneous and occur because of the inherent chemical reactivity of the DNA and the presence of normal, reactive chemical species within the cell. For example, the base cytosine undergoes spontaneous hydrolytic deamination to give uracil (*Fig.*).

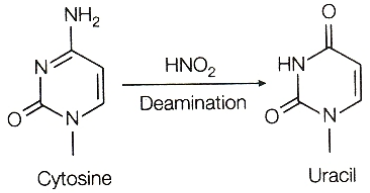


Fig. Deamination of cytosine to uracil by nitrous acid

If left unrepaired, the resulting uracil would form a base pair with adenine during subsequent replication, giving rise to a point mutation. In fact, the generation of uracil in DNA in this way is the probable reason why DNA contains thymine instead of uracil. Any uracil found in DNA is removed by an enzyme called uracil DNA glycosylase and is replaced by cytosine. 5-Methylcytosine, a modified base found in small amounts in DNA, deaminates to thymine, a normal base. This is much more difficult to detect.

Depurination is another spontaneous hydrolytic reaction that involves cleavage of the N-glycosylic bond between *N-9* of the purine bases A and G and C-l' of the deoxyribose sugar and hence loss of purine bases from the DNA. The sugar-phosphate backbone of the DNA

remains intact. The resulting apurinic site is a noncoding lesion, as information encoded in the purine bases is lost. Depurination occurs at the rate of 10000 purines lost per human cell per hour at 37°C. Though less frequent, depyrimidination can also occur.

Oxidative Damage: This occurs under normal conditions due to the presence of reactive oxygen species (ROO) in all aerobic cells, for example superoxide, hydrogen peroxide and, most importantly, the hydroxyl radical (-OH). This radical can attack DNA at a number of points, producing a range of oxidation products with altered properties, for example 8-oxoguanine, 2-oxoadenine and 5-formyluracil *(Fig).* The levels of these can be increased by hydroxyl radicals from the radiolysis of water caused by ionizing radiation.

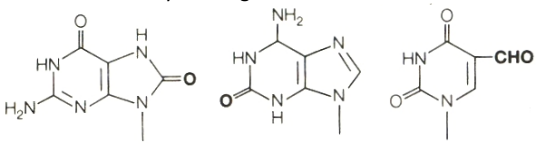


Fig. Examples of oxidized bases produced in DNA by reactive oxygen species.

Alkylation: Alkylating agents are electrophilic chemicals which readily add alkyl (e.g. methyl) groups to various positions on nucleic acids distinct from those methylated by normal methylating enzymes. Common examples are methylmethane sulfonate (MMS) and ethylnitrosourea (ENU) *(Fig).* Typical examples of methylated bases are 7-methylguanine, 3-methyl adenine, 3-methylguanine and O^6-methylguanine *(Fig.).* Some of these lesions are potentially lethal as they can interfere with the unwinding of DNA during replication and transcription. Most are also indirectly mutagenic; however, O^6-methylguanine is a directly mutagenic lesion as it can base-pair with thymine during replication.

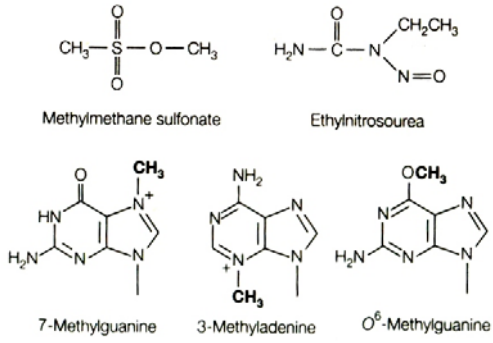
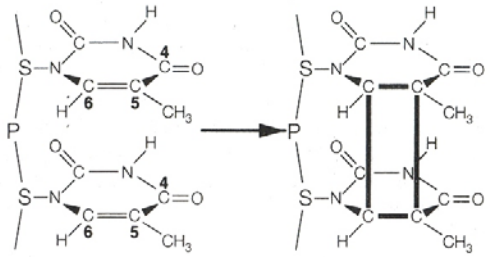


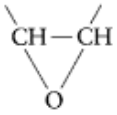
Fig. Examples of (a) alkylating agents; (b) alkylated bases.

Bulky Adducts: Cyclobutane pyrimidine dimers are formed by ultraviolet light from adjacent pyrimidines on one strand by cyc1ization of the double-bonded C5 and C6 carbon atoms of each base to give a cyclobutane ring *(Fig.).*

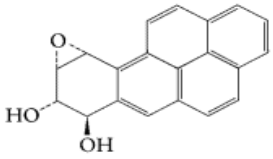
The resulting loss of base pairing with the opposite strand causes localized denaturation of the DNA producing a bulky lesion which would disrupt replication and transcription. Another type of pyrimidine dimer, the 6,4-photoproduct, results from the formation of a bond between C6 of one pyrimidine base and C4 of the adjacent base (see ring carbon numbers in *Fig).*



Detoxification begins with a powerful series of oxidation reactions catalyzed by a set of proteins called ***cytochrome P-450***. These enzymes, which are bound to endoplasmic reticulum membranes, can oxidize even highly unreactive compounds such as polycyclic aromatic hydrocarbons. Oxidation of polycyclic aromatics produces an epoxide, a very reactive electrophilic group:



Usually these epoxides are rapidly hydrolyzed into hydroxyl groups, which are then coupled to glucuronic acid or other groups, producing compounds soluble enough in water to be excreted. Some intermediate epoxides, however, are only slowly hydrolyzed to hydroxyl groups, probably because the relevant enzyme (epoxide hydratase) cannot get to the epoxide to act on it. For example, the indirect-acting carcinogen **benzo*(a)*pyrene**, undergoes two epoxidation reactions to yield a highly reactive electrophilic ultimate carcinogen:

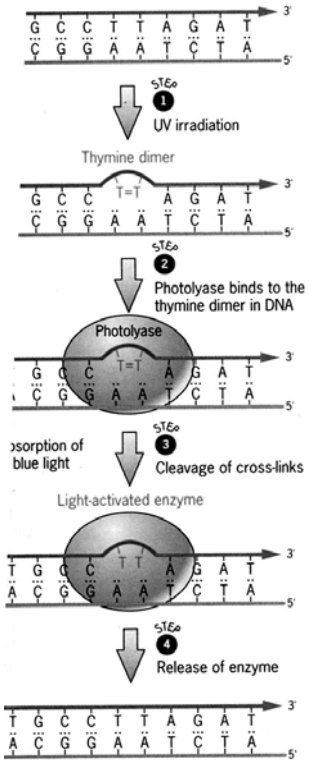


Other types of indirect-acting carcinogens are activated by different oxidative pathways, which also involve P-450 enzymes. The liver carcinogen **aflatoxin B_1** also covalently binds to DNA.

A4. DNA REPAIR MECHANISMS:

Errors during DNA replication and DNA damage due to exposure to mutagens happen all the time and microorganisms have evolved very complex systems to recognize and repair the damage. In the best situation the repair directly undoes the original damage, causing no change to the base sequence, but sometimes the amount of damage is so great that DNA synthesis cannot continue. In addition to the proofreading activity of DNA polymerases that can correct miscopied bases during replication, cells have evolved mechanisms for repairing DNA damaged by chemicals or radiation. Complex organisms with large genomes and relatively long generation times contain many cells that divide very slowly or not at all (e.g., liver and brain cells). Such cells must use the information in their DNA for weeks, months, or even years, greatly increasing their chances for sustaining damage to their DNA. If repair processes were 100 percent effective, chemicals and radiation would pose no threat to cellular DNA. Unfortunately, repair of lesions caused by some environmental agents is relatively inefficient, and such lesions can lead to mutations that ultimately cause cancer. In theory, a carcinogen could act by binding to DNA and causing a change in the sequence that is perpetuated during DNA replication. Current evidence suggests, however, that many permanent DNA sequence changes are induced by the very repair processes cells use to rid themselves of DNA damage.

4. 1. Direct Reversal:



a. Photo-reactivation: Cyclobutane pyrimidine dimers can be monomerized again by DNA **photolyases** (photo-reactivating enzymes) in the presence of visible light. These enzymes have prosthetic groups which absorb blue light and transfer the energy the cyclobutane ring which is then cleaved. The E. *coli* photo lyase has two chromophores, N^5,N^{10}-methenyltetrahydrofolate and reduced flavin adenine dinuotide (FADH). Photoreactivation is specific for pyrimidine dimers. It is an example of **direct reversal** of a lesion and is error-free.

b. Alkyl transferase: Another example of error-free direct reversal forms part of the **adaptive response** to alkylating agents. This response is induced in E. *coli* by low levels of alkylating agents and gives increased protection against the lethal and mutagenic effects of subsequent high doses. Mutagenic protection is afforded by an **alkyl transferase** which removes the alkyl group from the directly mutagenic O^6-alkylguanine. Curiously, the alkyl group is transferred to the protein itself and inactivates it. Thus, each alkyltransferase can only be used once. This protein is also present in mammalian cells. Protection against lethality involves induction of a DNA glycosylase which removes other alkylated bases through base excision repair.

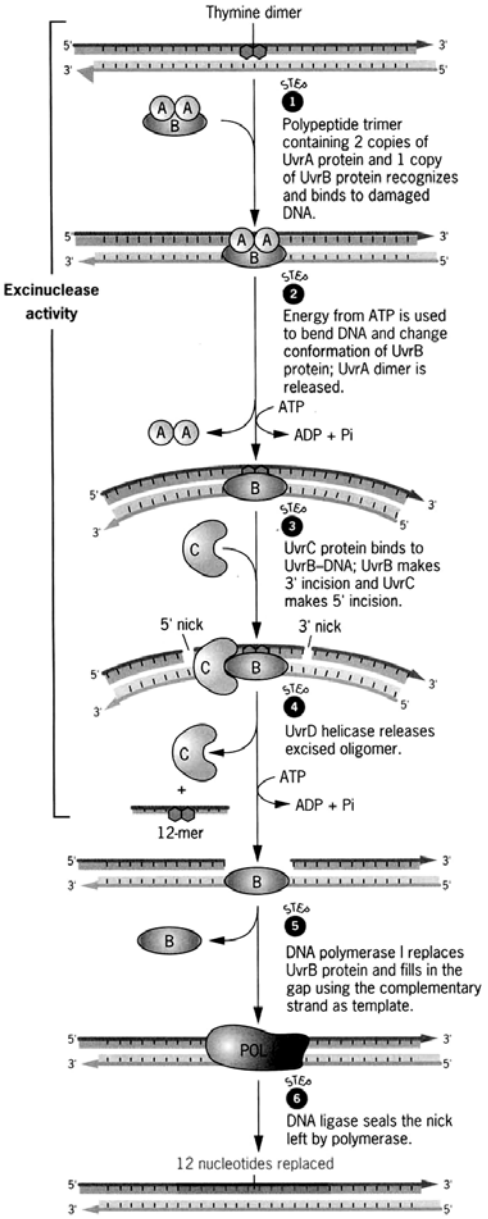
4. 2. Excision Repair Mechanism:

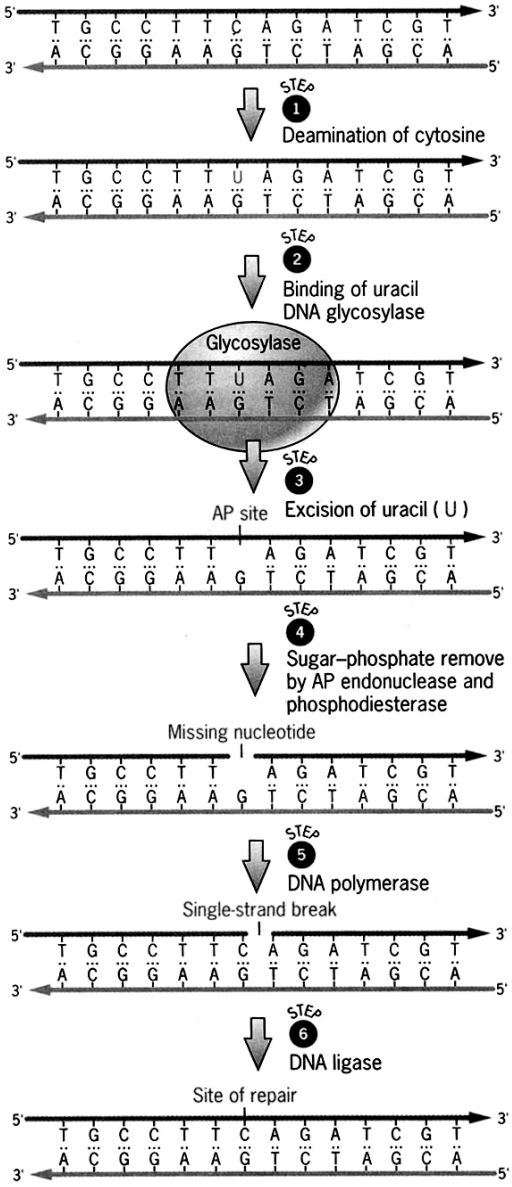
This ubiquitous mechanism operates on a wide variety of lesions and is essentially error-free. There are two forms, **Nucleotide excision repair (NER) and Base excision repair (BER)**.

In NER, an endonuclease cleaves the DNA a precise number of bases on either side of the lesion and an oligonucleotide containing the lesion is removed leaving a gap.

For example, in E. *coli,* the UvrABC endonuclease removes pyrimidine dimers and other bulky lesions by recognizing the distortion these produce in the double helix. In BER, modified bases are recognized by relatively specific DNA glycosylases which cleave the N-glycosylic bond between the altered base and the sugar CU, leaving an apurinic or apyrimidinic (AP) site. AP sites are also produced by spontaneous base loss. An AP endonuclease then cleaves the DNA at this site and a gap may be created by further exonuclease activity. The gap is generally larger in NER and can be as small as one nucleotide in BER. From this point on, both forms of excision repair are essentially the same. In E. *coli,* the gap is filled by DNA polymerase I and the final phospho-diester bond made by DNA ligase, much as in the final stages of processing of the lagging strand fragments during DNA replication. In eukaryotes, gap filling in BER involves predominantly DNA polymerase β whereas the

longer gaps generated in NER are filled by DNA polymerases **δ or ε** . In eukaryotic NER, recognition and excision of DNA damage is a complex process "involving at least 18 polypeptide factors, including the transcription factor TFIIH. Excision is coupled to transcription so that transcribed (genetically active) regions of the DNA are repaired more rapidly than non-transcribed DNA. This helps to limit the production of defective gene products.





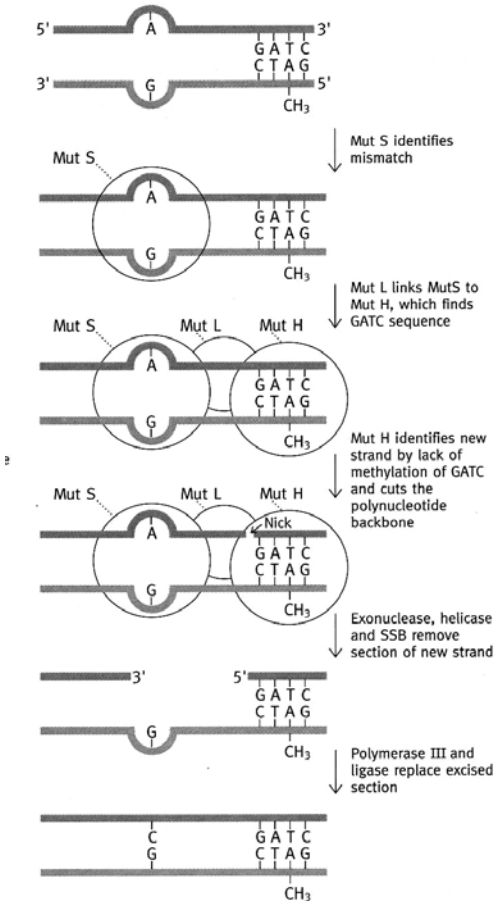
4. 3. Mismatch Repair:

Many spontaneous mutations are point mutations, which involve a change in a single base pair in the DNA sequence. These can arise from errors in replication, during genetic recombination, and, particularly, by *base deamination* whereby a C residue is converted into a U residue. Bacterial and eukaryotic cells have a mismatch-repair system that recognizes and repairs all single-base mispairs except C·C, as well as small insertions and deletions. The conceptual problem with mismatch repair is determining which is the normal and which is the mutant DNA strand, and repairing the latter so that it is properly base-paired with the normal strand.

In *E. coli* DNA, adenine residues in a GATC sequence are methylated at the 6 position. Since DNA polymerases incorporate adenine, not methyl-adenine, into DNA, adenine residues in newly replicated DNA are methylated only on the parental strand. The adenines in

GATC sequences on the daughter strands are methylated by a specific enzyme, called *Dam methyltransferase,* only after a lag of several minutes. During this lag period, the newly replicated DNA contains hemimethylated GATC sequences.

An *E. coli* protein designated *MutH,* which binds specifically to hemimethylated sequences, is able to distinguish the methylated parental strand from the unmethylated daughter strand. If an error occurs during DNA replication, resulting in a mismatched base pair near a GATC sequence, another protein, MutS, binds to this abnormally paired segment. Binding of MutS triggers binding of MutL, a linking protein that connects MutS with a nearby MutH. This cross-linking activates a latent endonuclease activity of MutH, which then cleaves specifically the unmethylated daughter strand. Following this initial incision, the segment of the daughter strand containing the misincorporated base is excised and replaced with the correct DNA sequence.



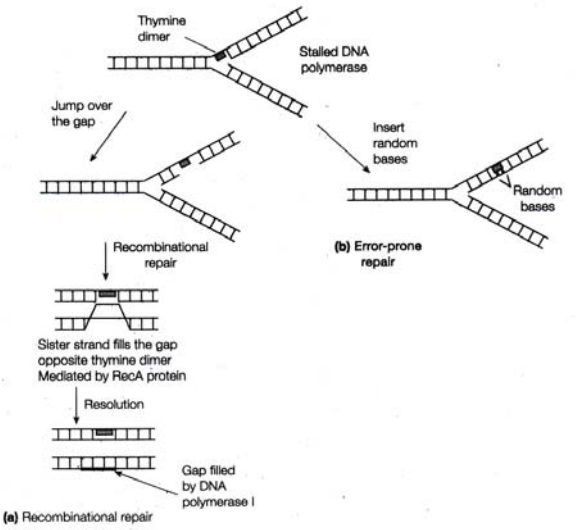
E. coli strains that lack the MutS, MutH, or MutL protein have a higher rate of spontaneous mutations than wild-type cells. Strains that cannot synthesize the Dam methyltransferase also have a high rate of spontaneous mutations. Because Dam strains cannot methylate adenines within GATC sequences, the MutHLS mismatch-repair system cannot distinguish between the template and newly synthesized strand and therefore cannot efficiently repair mismatched bases.

A similar mechanism repairs lesions resulting from *depurination,* the loss of a guanine or adenine base from DNA resulting from cleavage of the glycosidic bond between deoxyribose and the base. Depurination occurs spontaneously and is fairly common in mammals. The resulting *apurinic sites,* if left unrepaired, generate mutations during DNA replication because they cannot specify the appropriate paired base. All cells possess apurinic (AP) endonucleases that cut a DNA strand near an apurinic site. As with mismatch repair, the cut is extended by exonucleases, and the resulting gap then is repaired by DNA polymerase and ligase.

The discriminatory mechanism in eukaryotes is not known, but mismatch repair is clearly important. in maintaining the overall error rate of DNA replication and, therefore, the spontaneous mutation rate: hereditary non polyposis carcinoma of the colon is caused by mutational loss of one of the human mismatch repair enzymes. Mismatch repair may also correct errors that arise from sequence misalignments during meiotic recombination in eukaryotes.

4.4. Recombinational (Post Replication) Repair Mechanism:

Thymine dimers cannot be replicated but, instead of stalling at these sites, DNA polymerase can jump over the lesion and resume DNA synthesis further down the DNA template. This leaves a gap in the newly synthesized DNA strand opposite the dimer. This cannot be repaired by excision repair as this requires an intact strand to be used as the template. Instead the gap can be filled by Rec A-mediated recombination with the sister DNA helix, which contains the intact sequence that should be opposite the dimer *(Fig.).* Although recombination does not repair the damage it does create the situation where both new DNA molecules can be repaired by excision repair.



(a) Recombinational and (b) error-prone repair of UV-damaged DNA during replication.

4.5. Inducible DNA-Repair Systems (SOS Repair) Are Error-Prone

When a cell suffers so much DNA damage over a short time that its repair systems are saturated, it runs the danger of extensively replicating unrepaired lesions, thereby perpetuating mutations. In such situations, both bacterial and animal cells use inducible repair systems in an attempt to catch up. Such systems are not expressed in undamaged cells, but some aspect of the accumulated damage causes their derepression (induction) and expression.

There are a number of genes and operons in the cell that are coordinately regulated by RecA protein and a repressor of transcription, LexA. These are involved in dealing with DNA damage and are called SOS-repair systems. These include an error-prone repair system which interacts with DNA polymerase allowing it to continue replication of DNA past a pyrimidine dimer. The 3'→5' proof-reading ability of the enzyme is inhibited so that random bases can be inserted opposite the dimer without accurate base-pairing. This mechanism is the main reason why UV light is mutagenic because most of the other systems repair DNA accurately.

The SOS-repair systems are induced by RecA protein, activated by a conformation change, caused by the presence of damaged DNA. The activated RecA proteins causes LexA protein to be proteolytically cleaved so that it no longer acts as a repressor of transcription. The genes in the system may then be expressed as long as the DNA damage is present in the cell. Once the damage is removed, RecA protein is no longer activated as an inducer of LexA proteolysis, so LexA protein accumulates in the cell and inhibits transcription of the SOS-repair genes.

Animal cells also have inducible repair systems, although it is not known whether these are error-prone. As noted earlier, however, the main mechanism for repairing double-strand breaks in eukaryotes clearly is error-prone. Thus, double-strand repair and, perhaps, error-prone inducible repair likely play a role in mutagenesis and therefore in carcinogenesis in animals. In any case, many investigators believe that in animal cells, as in bacteria, most mutations are an indirect, not direct, consequence of DNA damage.

Because radiation- or carcinogen-induced DNA damage must be repaired before the DNA is replicated, cells have sensing mechanisms that react to DNA damage and stop DNA replication. These mechanisms, which are discussed in detail in the next chapter, involve checkpoint control proteins such as the p53 protein, which acts to stop the cell cycle if DNA is damaged, and thus to suppress production of tumors. Cells that do not express functional p53 protein exhibit high rates of mutation in response to DNA damage, accelerating the formation of tumors.

Xeroderma pigmentosum (XP) is an autosomal recessive disorder characterized phenotypically by extreme sensitivity to sunlight and a high incidence of skin tumors. XP sufferers are defective in the NER of bulky DNA damage, including that caused by ultraviolet light. Defects in at least seven different genes can cause XP, indicating the complexity of excision repair in mammalian cells. Xeroderma pigmentosum variant (XP-V) is clinically very similar to classical XP but cells from XP-V individuals can carry out normal NER. In this case, the defect is in the gene encoding the translesion DNA polymerase η, which normally inserts dAMP in an error-free fashion opposite the thymine residues in a cyclobutane thymine dimer. XP-V cells may, therefore, have to rely more heavily on alternative error-prone modes of translesion DNA synthesis to maintain DNA integrity after radiation damage. Sufferers of **Cockayne syndrome** are also sun-sensitive and defective in transcription coupled excision repair, but are not cancer-prone.

B1. BASIC PRINCIPLES OF TRANSCRIPTION

Transcription is the enzymic synthesis of RNA on a DNA template. This is the first stage in the overall process of gene expression and ultimately leads to synthesis of the protein encoded by a gene. Transcription is catalyzed by an RNA polymerase which requires a dsDNA template as well as the precursor ribonucleotides ATP, GTP, err and UTP ***(Fig.1***). RNA synthesis always occurs in a fixed direction, from the 5'- to the 3'-end of the RNA molecule. Usually, only one of the two strands of DNA becomes transcribed into RNA. One strand is known as the sense strand. The sequence of the RNA is a direct copy of the sequence of the deoxynucleotides in the **sense strand** (with U in place of T). The other strand is known as the **antisense strand**. This strand may also be called the template strand since it is used as the template to which ribonucleotides base-pair for the synthesis of the RNA.

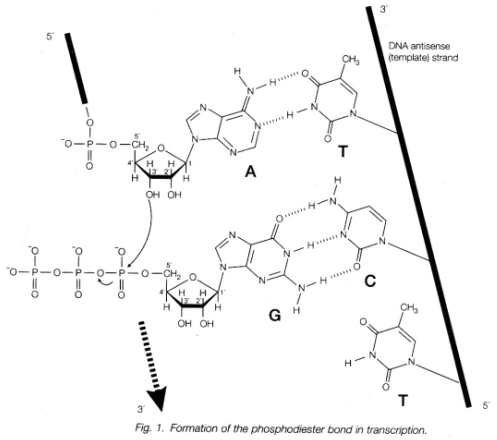


Fig. 1. Formation of the phosphodiester bond in transcription.

1.1 Initiation of Transcription: Initiation of transcription involves the binding of an RNA polymerase to the dsDNA. RNA polymerases are usually multisubunit enzymes. They bind to the dsDNA and initiate transcription at sites called **promoters** ***(Fig.* 2)**. Promoters are sequences of DNA at the start of genes, that is to the 5' -side (**upstream**) of the coding region. Sequence elements of promoters are often conserved between different genes. Differences between the promoters of different genes give rise to differing efficiencies of transcription initiation and are involved in their regulation. The short conserved sequences within promoters are the sites at which the polymerase or other DNA-binding proteins bind to initiate or regulate transcription.

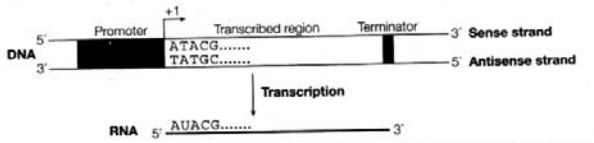


Fig. 2. Structure of a typical transcription unit showing promoter and terminator sequence and the RNA product.

In order to allow the template strand to be used for base pairing, the DNA helix must be locally unwound. Unwinding begins at the promoter site to which the RNA polymerase binds. The polymerase then initiates the synthesis of the RNA strand at a specific nucleotide called the start site (**initiation site**). This is defined as position +1 of the gene sequence ***(Fig.* 2)**. The RNA polymerase and its co-factors, when assembled on the DNA template, are often referred to as the **transcription complex**.

1.2 Elongation: The RNA polymerase covalently adds ribonucleotides to the 3'-end of the growing RNA chain *(Fig.* 1). The polymerase therefore extends the growing RNA chain in a 5' →3' direction. This occurs while the enzyme itself moves in a 3' →5' direction along the antisense DNA strand (template). As the enzyme moves, it locally unwinds the DNA, separating the DNA strands, to expose the template strand for ribonucleotide base pairing and covalent addition to the 3'end of the growing RNA chain. The helix is reformed behind the polymerase. The E. *coli* RNA polymerase performs this reaction at a rate of around 40 bases per second at 37°C.

1.3 Termination: The termination of transcription, namely the dissociation of the transcription complex and the ending of RNA synthesis, occurs at a specific DNA sequence known as the **terminator**. These sequences often contain self-complementary regions which can form a **stem-loop or hairpin** secondary structure in the RNA product ***(Fig.* 3)**. These cause the polymerase to pause and subsequently cease transcription.

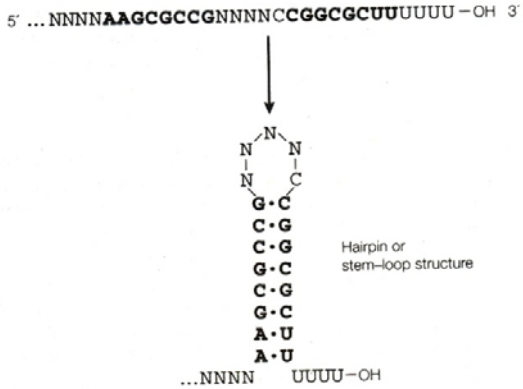


Fig. 3. RNA hairpin structure.

Some terminator sequences can terminate transcription, without the requirement for accessory factors, whereas other terminator sequences require the **rho protein** (ρ) as an accessory factor. In the termination reaction, the RNA-DNA hybrid is separated allowing the reformation of the dsDNA, and the RNA polymerase and synthesized RNA are released from the DNA.

B2. TRANSCRIPTION IN PROKARYOTES (*E .coli*)

2.1 *Escherichia Coli* RNA Polymerase

The E. *coli* RNA polymerase is one of the largest enzymes in the cell. The enzyme consists of at least five subunits. These are the alpha (α), beta (β), beta prime (β′), omega (*w*) and sigma (σ) subunits. In the complete polymerase called the holoenzyme, there are two a subunits and one each of the other four subunits (i.e. $\alpha_2\beta\beta'\omega\sigma$). The complete enzyme is required for transcription initiation. However, the σ factor is not required for transcription elongation and is released from the transcription complex after transcription initiation. The remaining enzyme, which translocates along the DNA, is known as the core enzyme and has the structure $\alpha_2\beta\beta'\omega$. The E. *coli* RNA polymerase can synthesize RNA at a rate of around 40 nucleotide per sec at 37°C and requires Mg^{2+} for its activity. The enzyme has a non-spherical structure with a projection flanking a cylindrical channel. The size of the channel suggests that it can bind directly to 16 bp of DNA. The whole polymerase binds over a region of DNA covering around 60 bp.

Although most RNA polymerases like the E. *coli* polymerase have a multisubunit structure, it is important to note that this is not an absolute requirement. The RNA polymerases encoded by bacteriophages T3 and T7 are single polypeptide chains which are much smaller than the bacterial multisubunit enzymes. They synthesize RNA rapidly (200 nt per see at 37°C) and recognize their own specific DNA-binding sequences.

α Subunit: Two identical α subunits are present in the core RNA polymerase enzyme. The subunit is encoded by the *rpoA* gene. The α subunit is required for core protein assembly, but has had no clear transcriptional role assigned to it. When phage T4 infects *E. coli* the α subunit is modified by adenosine diphosphate (ADP) ribosylation of an arginine. This is associated with a reduced affinity for binding to promoters, suggesting that the α subunit may play a role in promoter recognition.

β Subunit: One β subunit is present in the core enzyme. This subunit is thought to be the catalytic center of the RNA polymerase. Strong evidence for this has come from studies with antibiotics which inhibit transcription by RNA polymerase. The important antibiotic **rifampicin** is a potent inhibitor of RNA polymerase that blocks initiation but not elongation. This class of antibiotic does not inhibit eukaryotic polymerases and has, therefore, been used medically for treatment of Gram-positive bacteria infections and tuberculosis. Rifampicin has been shown to bind to the β subunit. Mutations that give rise to resistance to rifampicin map to *rpoB,* the gene that encodes the β subunit. A further class of antibiotic, the **streptolydigins**, inhibit transcription elongation, and mutations that confer resistance to these antibiotics also map to *rpoB.* These studies suggest that the β subunit may contain two domains responsible for transcription initiation and elongation.

β' Subunit: One β' subunit is present in the core enzyme. It is encoded by the *rpoC* gene. This subunit binds two Mg^{2+} ions which are thought to participate in the catalytic function of the polymerase. A polyanion, **heparin**, has been shown to bind to the β' subunit. Heparin inhibits transcription *in vitro* and also competes with DNA for binding to the polymerase. This suggests that the β' subunit may be responsible for binding to the template DNA.

Sigma factor: The most common sigma factor in E. *coli* is σ^{70} (since it has a molecular mass of 70 kDa). Binding of the σ factor converts the core RNA polymerase enzyme into the holoenzyme. The σ factor has a critical role in promoter recognition, but is not required for transcription elongation. The σ factor contributes to promoter recognition by decreasing the affinity of the core enzyme for nonspecific DNA sites by a factor of 104 and increasing affinity for the promoter. Many prokaryotes (including E. *coli)* have multiple σ factors. They are involved in the recognition of specific classes of promoter sequences. The σ factor is released from the RNA polymerase when the RNA chain reaches 8-9 nt in length. The core enzyme then moves along the DNA synthesizing the growing RNA strand. The σ factor can then complex with a further core enzyme complex and re-initiate transcription. There is only 30% of the amount of σ factor present in the cell compared with core enzyme complexes. Therefore only one-third of the polymerase complexes can exist as holoenzyme at anyone time.

2.2 The *E. coli* σ^{70} Promoter:

RNA polymerase binds to specific initiation sites upstream from transcribed sequence. These are called promoters. Although different promoters are recognized by different a factors which interact with the RNA polymerase core enzyme, the most common σ factor in E. *coli* is σ^{70}. Promoters were first characterized through mutations that enhance or diminish the rate of transcription of genes such as those in the *lac* operon. The promoter lies upstream of the start site of transcription, generally assigned as position +1. In accordance with this, promoter sequences are assigned a negative number reflecting the distance upstream from the start of transcription. Mutagenesis of E. *coli* promoters has shown that only very short conserved sequences are critical for promoter function.

Promoter Size: The σ^{70} promoter consists of a sequence of between 40 and 60 bp. The region from around -55 to +20 has been shown to be bound by the polymerase, and the region from -20 to +20 is strongly protected from nuclease digestion by DNase I. This suggests that this region is tightly associated with the polymerase which blocks access of the nuclease to the DNA. Mutagenesis of promoter sequences showed that sequences up to around position -40 are critical for promoter function. Two 6 bp sequences at around positions -10 and -35 have been shown to be particularly important for promoter function in E. *coli.*

-10 Sequence: The most conserved sequence in σ^{70} promoters is a 6 bp sequence which is found in the promoters of many different E. *coli* genes. This sequence is centered at around the -10 position with respect to the transcription start site ***(Fig.* 4)**. This is sometimes referred to as the **Pribnow box**, having been first recognized by Pribnow in 1975. It has a consensus sequence of TATAAT, where the consensus sequence is made up of the most frequently occurring nucleotide at each position when many sequences are compared. The first two bases (TA) and the final T are most highly conserved. This hexamer is separated by between 5 and 8 bp from the transcription start site. This intervening sequence is not conserved, although the distance is critical. The -10 sequence appears to be the sequence at which DNA unwinding is initiated by the polymerase.

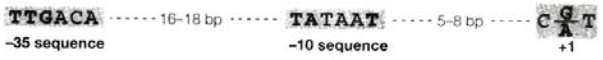


Fig. 4 *Consensus sequences of* E. coli *promoters*

-35 Sequence: Upstream regions around position -35 also have a conserved hexamer sequence **(*Fig.* 4)**. This has a consensus sequence of TTGACA, which is most conserved in efficient promoters. The first three positions of this hexamer are the most conserved. This sequence is separated by 16-18 bp from the -10 box in 90% of all promoters. The intervening sequence between these conserved elements is not important.

Transcription Start Site: The transcription start site is a purine in 90% of all genes ***(Fig.* 4)**. G is more common at the transcription start site than A. Often, there are C and T bases on either side of the start site nucleotide (i.e. CGT or CAT).

Promoter Efficiency: The sequences described above are consensus sequences typical of strong promoters. However, there is considerable variation in sequence between different promoters, and they may vary in transcriptional efficiency by up to 1000-fold. Overall, the functions of different promoter regions can be defined as follows:

- The -35 sequence constitutes a recognition region which enhances recognition and interaction with the polymerase σ factor;
- The -10 region is important for DNA unwinding;
- The sequence around the start site influences initiation.

The sequence of the first 30 bases to be transcribed also influences transcription. This sequence controls the rate at which the RNA polymerase clears the promoter, allowing re-initiation of another polymerase complex, thus influencing the rate of transcription and hence the overall promoter strength. The importance of strand separation in the initiation reaction is shown by the effect of negative supercoiling of the DNA template which generally enhances transcription initiation, presumably because the supercoiled structure requires less energy to unwind the DNA. Some promoter sequences are not sufficiently similar to the consensus sequence to be strongly transcribed under normal conditions. An example is the ***lac* promoter P*lac*,** which requires an accessory activating factor called **cAMP receptor protein (CRP)** to bind to a site on the DNA close to the promoter sequence in order to enhance polymerase binding and transcription initiation. Other promoters, such as those of genes associated with heat shock, contain different consensus promoter sequences that can only be recognized by an RNA polymerase which is bound to σ factor different from the general factor σ^{70}.

2.3. Initiation, Elongation and Termination of Transcription:

Promoter Binding: The RNA polymerase core enzyme $\alpha_2\beta\beta'\omega$, has a general nonspecific affinity for DNA. This is referred to as **loose binding** and it is fairly stable. When σ factor is added to the core enzyme to form the holoenzyme, it markedly reduces the affinity for nonspecific sites on DNA by 20,000-fold. In addition, σ factor enhances holoenzyme binding to correct promoter-binding sites 100 times. Overall, this dramatically increases the specificity of the holoenzyme for correct promoter-binding sites. The holoenzyme searches out and binds to promoters in the E. *coli* genome extremely rapidly. This process is too fast to be achieved by repeated binding and dissociation from DNA, and is believed to occur by I the polymerase sliding along the DNA until it reaches the promoter sequence. At the promoter, the polymerase recognizes the double-stranded -35 and -10 DNA sequences. The initial complex of the polymerase with the base-paired promoter DNA is referred to as a **closed complex**.

DNA Unwinding: In order for the antisense strand to become accessible for base pairing, the DNA duplex must be unwound by the polymerase. Negative supercoiling enhances the transcription of many genes, since this facilitates unwinding by the polymerase. However, some promoters are not activated by negative supercoiling, implying that differences in the natural DNA topology may affect transcription, perhaps due to differences in the steric relationship of the -35 and -10 sequences in the double helix. For example, the promoters for the enzyme subunits of **DNA gyrase** are inhibited by negative supercoiling. DNA gyrase is responsible for negative supercoiling of the E. *coli* genome and so this may serve as an elegant feedback loop for DNA gyrase protein expression. The initial unwinding of the DNA results in formation of an **open complex** with the polymerase; and this process is referred to as **tight binding**.

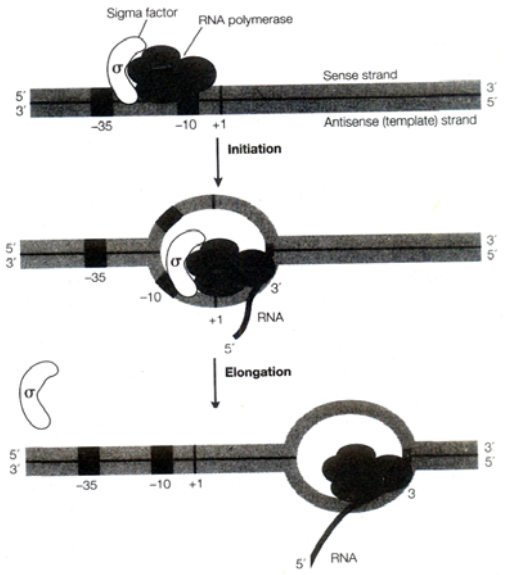


Fig. 5 Formation of the transcription complex: initiation and elongation.

RNA Chain initiation: Almost all RNA start sites consist of a purine residue, with G being more common than A. Unlike DNA synthesis, RNA synthesis can occur without a primer ***(Fig.* 5)**. The chain is started with a GTP or ATP, from which synthesis of the rest of the chain is initiated. The polymerase initially incorporates the first two nucleotides and forms a phosphodiester bond between them. The first nine bases are added without enzyme movement along the DNA. After each one of these first 9 nt is added to the chain, there is a significant probability that the chain will be aborted. This process of abortive initiation is important for the overall rate of transcription since it has a major role in determining how long the polymerase takes to leave the promoter

and allow another polymerase to initiate a further round of transcription. The minimum time for promoter clearance is 1-2 seconds, which is a long event relative to other stages of transcription.

RNA Chain elongation: When initiation succeeds, the enzyme releases the α factor and forms a ternary complex (three components) of polymerase-DNA-nascent (newly synthesized) RNA, causing the polymerase to progress along the DNA (promoter clearance) allowing re-initiation of transcription from the promoter by a further RNA polymerase holoenzyme.

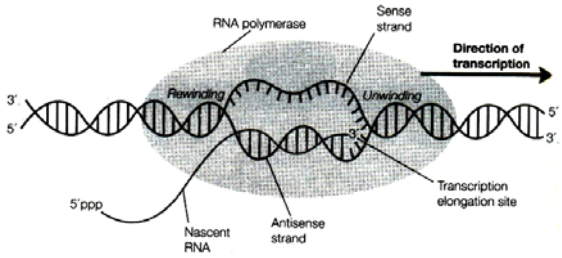


Fig. 6. Schematic structure of the transcription bubble during elongation.

The region of unwound DNA, which is called the transcription bubble, appears to move along the DNA with the polymerase. The size of this region of unwound DNA remains constant at around 17bp ***(Fig.* 6)**, and the 5'-end of the RNA forms a hybrid helix of about 12 bp with the antisense DNA strand. This corresponds to just less than one turn of the RNA-DNA helix. The E. *coli* polymerase moves at an average rate of 40 nt per sec, but the rate can vary depending on local DNA sequence. Maintenance of the short region of unwound DNA indicates that the polymerase unwinds DNA in front of the transcription bubble and rewinds DNA at its rear. The RNA-DNA helix must rotate each time a nucleotide is added to the RNA.

RNA Chain termination: The RNA polymerase remains bound to the DNA and continues transcription until it reaches a **terminator sequence (stop signal)** at the end of the transcription unit ***(Fig.* 7)**. The most common stop signal is an RNA hairpin in which the RNA transcript is self-complementary. As a result, the RNA can form a stable hairpin structure with a stem and a loop. Commonly the stem structure is very GC-rich, favoring its base pairing stability due to the additional stability of G-C base pairs over *A-U* base pairs. The RNA hairpin is often followed by a sequence of four or more U residues. It seems that the polymerase pauses immediately after it has synthesized the hairpin RNA. The subsequent stretch of U residues in the RNA base-pairs only weakly with the corresponding A residues in the antisense DNA strand. This favors dissociation of the RNA from the complex with the template strand of the DNA. The RNA is therefore released from the transcription complex. The non-base-paired antisense strand of the DNA then re-anneals with the sense DNA strand and the core enzyme disassociates from the DNA.

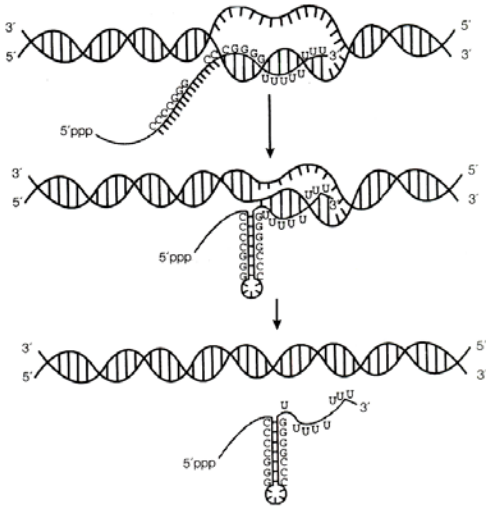


Fig. 7 Schematic diagram of rho-independent transcription termination.

Rho-dependent termination:

While the RNA polymerase can self-terminate at a hairpin structure followed by a stretch of U residues, other known terminator sites may not form strong hairpins. They use an accessory factor, the **rho protein (*p*)** to mediate transcription termination. Rho is a hexameric protein that hydrolyzes ATP in the presence of single-stranded RNA. The protein appears to bind to a stretch of 72 nucleotides in RNA, probably through recognition of a specific structural feature rather than a consensus sequence. Rho moves along the nascent RNA towards the transcription complex. There, it enables the RNA polymerase to terminate at rho-dependent transcriptional terminators. Like rho-independent terminators, these signals are recognized in the newly synthesized RNA rather than in the template DNA. Sometimes, the rho-dependent terminators are hairpin structures which lack the subsequent stretch of U residues which are required for rho-independent termination.

B3. TRANSCRIPTION IN EUKARYOTES

The mechanism of eukaryotic transcription is similar to that in prokaryotes. However, the large number of polypeptides associated with the eukaryotic transcription machinery makes it far more complex.

3.1 Eukaryotic RNA Polymerases: Three different RNA polymerase complexes are responsible for the transcription of different types of eukaryotic genes. The different RNA polymerases were identified by **chromatographic purification** of the enzymes and elution at different salt concentrations. Each RNA polymerase has a different sensitivity to the fungal toxin α-amanitin and this can be used to distinguish their activities.

- RNA polymerase I (RNA Pol I) transcribes most rRNA genes. It is located in the nucleoli and is insensitive to α-amanitin
- RNA polymerase II (RNA Pol II) transcribes all protein-coding genes and some small nuclear RNA (snRNA) genes. It is located in the nucleoplasm and is very sensitive to α-amanitin.
- RNA polymerase III (RNA Pol III) transcribes the genes for tRNA, 5 S rRNA, U6 snRNA and certain other small RNAs. It is located in the nucleoplasm and is moderately sensitive to α-amanitin.

In addition to these nuclear enzymes, eukaryotic cells contain additional polymerases in mitochondria and chloroplasts.

RNA polymerase Subunits: All three polymerases are large enzymes containing 12 or more subunits. The genes encoding the two largest subunits of' each RNA polymerase have homology (related DNA coding sequences) to each other. All of the three eukaryotic polymerases contain subunits which have homology to subunits withiri the E. *coli* core RNA polymerase $\alpha_2\beta\beta'$. The largest subunit of each eukaryotic RNA polymerase is similar to the β' subunit of the *E.* coli polymerase, and the second largest subunit is similar to the β subunit which contains the active site of the E. *coli* enzyme. The functional significance of this homology is supported by the observation that the second largest subunits of the eukaryotic RNA polymerases also contain the active sites. Two subunits which are common to RNA Pol I and RNA Pol III, and a further subunit which is specific to RNA Pol II, have homology to the E. *coli* RNA polymerase a subunit. At least five other smaller subunits are common to the three different polymerases. Each polymerase also contains an additional four to seven subunits which are only present in one type.

Eukaryotic RNA polymerase Activity: Like bacterial RNA polymerases, each of the eukaryotic enzymes catalyzes transcription in a 5' to 3' direction and synthesizes RNA complementary to the antisense template strand. The reaction requires the precursor nucleotides ATP, GTP, CTP and UTP and does not require a primer for transcription initiation. The purified eukaryotic RNA polymerases, unlike the purified bacterial enzymes, require the presence of additional initiation proteins before they are able to bind to promoters and initiate transcription.

The CTD of RNA Pol II: The carboxyl end of RNA Pol II contains a stretch of seven amino acids that is repeated 52 times in the mouse enzyme and 26 times in yeast. This heptapeptide has the sequence **Tyr-Ser-Pro- Thr-Ser-Pro-Ser** and is known as the **carboxylterminal domain** or CTD. These repeats are essential for viability. The CTD sequence may be phosphorylated at the serines and some tyrosines. *In vitro* studies have shown that the CTD is unphosphorylated at transcription initiation, but phosphorylation occurs during transcription elongation as the RNA polymerase leaves the promoter. Since RNA Pol II catalyzes the synthesis of all of the eukaryotic protein-coding genes, it is the most important RNA polymerase for the study of differential gene expression. The CTD has been shown to be an important target for differential activation of transcription elongation

3.2 RNA Polymerase II Genes: Promoters and Enhancers:

RNA polymerase II (RNA Pol II) is located in the nucleoplasm. It is responsible for the transcription of all protein-coding genes and some small nuclear RNA genes. The pre-mRNAs must be processed after synthesis by cap formation at the 5'-end of the RNA and poly(A) addition at the 3'-end, as well as removal of introns by splicing.

Promoters: Many eukaryotic promoters contain a sequence called the **TATA box** around 25-35 bp upstream from the start site of transcription ***(Fig.* 8**). It has the 7 bp consensus sequence 5'-TATA(A/TIA(A/T)-3' although it is now known that the protein which binds to the TAT A box, TBP, binds to an 8 bp sequence that includes an additional downstream base pair, whose identity is not important. The TAT A box acts in a similar way to an E. *coli* promoter -10 sequence to position the RNA Pol II for correct transcription initiation. While the sequence around the TATA box is critical, the sequence between the TAT A box and the transcription start site is not critical. However, the spacing between the TATA box and the start site is important. Around 50% of the time, the start site of transcription is an adenine residue.

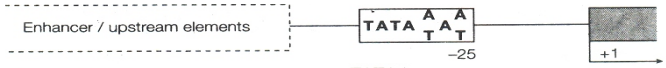


Fig. **8** *RNA Pol II promoter containing TATA box.*

Some eukaryotic genes contain an initiator element instead of a TATA box. The initiator element is located around the transcription start site. Many initiator elements have a C at position -1 and an A at +1. Other promoters have neither a TATA box nor an initiator element. These genes are generally transcribed at low rates, and initiation of transcription may occur at different start sites over a length of up to 200 bp. These genes often contain a GC-rich 20-50 bp region within the first 1-200 bp upstream from the start site.

Upstream regulatory elements: The low activity of basal promoters is greatly increased by the presence of other elements located upstream of the promoter. These elements are found in many genes which vary widely in their levels of expression in different tissues. Two common examples are the **SP1** box, which is found upstream of many genes both with and without **TATA boxes**, and the **CCAAT box**. Promoters may have one or both or multiple copies of these sequences. These sequences which are often located within 1-200 bp upstream from the promoter are referred to as **upstream regulatory elements** (UREs) and play an important role in ensuring efficient transcription from the promoter.

Enhancers: Transcription from many eukaryotic promoters can be stimulated by control elements that are located many thousands of base pairs away from the transcription start site. This was first observed in the genome of the DNA virus SV40. A sequence of around 100 bp from SV40 DNA can significantly increase transcription from a basal promoter even when it is placed far upstream. Enhancer sequences are characteristically 10-200 bp long and contain multiple sequence elements which contribute to the total activity or the enhancer. They may be ubiquitous or cell type-specific. Classically, enhancers have the following general characteristics:

- They exert strong activation of transcription of a linked gene from the correct start site.
- They activate transcription when placed in either orientation with respect to linked genes.
- They are able to function over long distances of more than 1 kb whether from an upstream or downstream position relative to the start site.
- They exert preferential stimulation of the closest of two tandem promoters.

However, as more enhancers and promoters have been identified, it has been shown that the upstream promoter and enhancer motifs overlap physically and functionally. There seems to be a continuum between classic enhancer elements and those promoter elements which are orientation specific and must be placed close to the promoter to have an effect on transcriptional activity.

3.3 General Transcription Factors for RNA Polymerase II (basal transcription factors):

A series of nuclear transcription factors have been identified, purified and cloned. These are required for basal transcription initiation from RNA Pol II promoter sequences *in vitro.* These multisubunit factors are named transcription factor IIA, 1m, etc. (TFIIA, etc.). They have been shown to assemble on basal promoters in a specific order *(Fig.* 9) and they may be subject to multiple levels of regulation.

TFII D: In promoters containing a TATA box, the RNA Pol II transcription factor TFIID is responsible for binding to this key promoter element. The .binding of TFIID to the TATA box is the earliest stage in the formation of the RNA Pol II transcription initiation complex. TFIID is a multiprotein complex in which only one polypeptide, **TATA-binding protein** (TBP) binds to the TATA box. The complex also contains other polypeptides known as **TBP-associated factors ($TAF_{II}s$)**. It seems that in mammalian cells, TBP binds to the TATA box and is then joined by at least eight $TAF_{II}s$ to form TFIID.

TBP: TBP is present in all three eukaryotic transcription complexes (in SL1, TFIIIB and TFIID) and clearly plays a major role in transcription initiation. TBP is a monomeric protein. All eukaryotic TBPs analyzed have very highly conserved C-terminal domains of 180 residues and this conserved domain functions as well as the full-length protein in *in vivo* transcription. The function of the less conserved N-terminal domain is therefore not fully understood. TBP has been shown to have a saddle structure with an overall dyad symmetry, but the two halves of the molecule are not identical. TBP interacts with DNA in the minor groove so that the inside of the saddle binds to DNA at the TATA box and the outside surface of the protein is available for interactions with other protein factors. Binding of TBP deforms the DNA so that it is bent into the inside of the saddle and unwound. This results in a kink of about 45^0 between the first two and last two base pairs of the 8 bp TATA element. A TBP with a mutation in its TAT A box-binding domain retains its function for transcription by RNA Pol I and Pol III, but it inhibits transcription initiation by RNA Pol II. This indicates that the other two polymerases use TBP to initiate transcription, but the precise role of TBP in these complexes remains unclear.

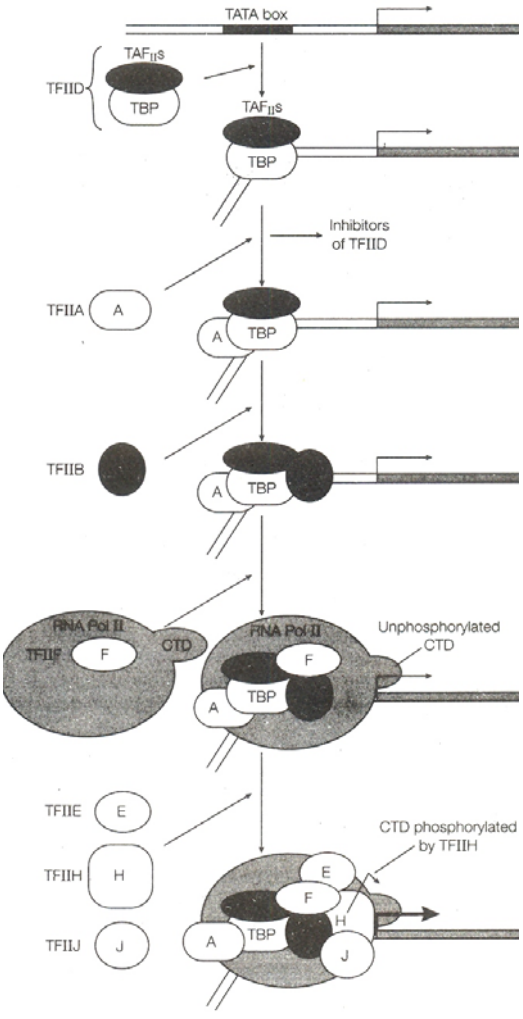


Fig. 9. A Schematic diagram of the assembly of the RNA pol II transcription initiation complex at a TATA box promoter

TFIIA: TFIIA binds to TFIID and enhances TFIID binding to the TATA box, stabilizing the TFIID-DNA complex. TFIIA is made up of at least three subunits. In *in vitro* transcription studies, as TFIID is purified, the requirement for TFIIA is lost. In the intact cell, TFIIA appears to counteract the effects of inhibitory factors such as DR1 and DR2 with which TFIID is associated. It seems likely that TFIIA binding to TFIID prevents binding of these inhibitors and allows the assembly process to continue.

TFIID and RNA polymerase binding: Once TFIID has bound to the DNA, another transcription factor, TFIIB, binds to TFIID. TFIIB can also bind to the RNA polymerase. This seems to be an important step in transcription initiation since TFIIB acts as a bridging factor allowing recruitment of the polymerase to the complex together with a further factor, **TFIIF**.

Factor binding RNA polymerase: After RNA polymerase binding, three other transcription factors, **TFIIE, TFIIH and TFIIJ**, rapidly associate with the complex. These proteins are necessary for transcription *in vitro* and associate with the complex in a defined order. **TFIIH** is a large complex which is made up of at least five subunits. TFIIJ remains to be fully characterized.

CTD phosphorylation by TFIIH: TFIIH is a large multi-component protein complex which contains both **kinase and helicase activity**. Activation of TFIIH results in **phosphorylation** of the carboxyl-terminal domain (CTD) of the RNA polymerase. This **phosphorylation** results in formation of a processive RNA polymerase complex and allows the RNA polymerase to leave the promoter region. TFIIH therefore seems to have a very important function in control of transcription elongation. Components of TFIIH are also important in DNA repair and in phosphorylation of the cyclin-dependent kinase complexes which regulate the cell cycle.

The Initiator transcription Complex: Many RNA Pol II promoters which do not contain a TATA box have an initiator element overlapping their start site. It seems that at these promoters TBP is recruited to the promoter by a further DNA-binding protein which binds to the initiator element. TBP then recruits the other transcription factors and RNA polymerase in a manner similar to that which occurs in TATA box promoters.

B4. m-RNA PROCESSING

There appears to be little or no processing of mRNA transcripts in prokaryotes. In fact, ribosomes can assemble on, and begin to translate, mRNA molecules that have not yet been completely synthesized. Prokaryotic mRNA is degraded rapidly from the 5'-end and the first cistron (protein-coding region) can therefore only be translated for a limited amount of time. Some internal cistrons are partially protected by stem-loop structures that form at the 5'- and 3'-ends and provide a temporary barrier to exonucleases and can thus be translated more often before they are eventually degraded. Because eukaryotic RNA Pol II transcribes such a wide variety of different genes, from the snRNA genes of 60-300 nt to the large *Antennapedia* gene, whose transcript can be over 100 kb in length, the collection of products made by this enzyme is referred to as heterogeneous nuclear RNA (hnRNA). Those transcripts that will be processed to give mRNAs are called pre-mRNAs. Pre-mRNA molecules are processed to mature mRNA by 5'-capping 3'-cleavage and polyadenylation, splicing and methylation.

hnRNP: The hnRNA synthesized by RNA Pol II is mainly pre-mRNA and rapidly becomes covered in proteins to form heterogeneous nuclear ribonucleoprotein (hnRNP). The proteins involved have been classified as hnRNP proteins A-U. There are two forms of each of the three more abundant hnRNP proteins, the A, Band C proteins. Purification of this material from nuclei gives a fairly homogeneous preparation of 30-405 particles called hnRNP particles. These particles are about 20 nm in diameter and each contains about 600-700 nt of RNA complexed with three copies of three different tetramers. These tetramers are $(A_1)_3B_2$, $(A_2)_3B_1$ and $(C_1)_3C_2$. The hnRNP proteins are thought to help keep the hnRNA in a single-stranded form and to assist in the various RNA processing reactions.

SnRNP particles: RNA Pol II also transcribes most snRNAs which complex with specific proteins to form **snRNPs**. These RNAs are rich in the base uracil and are thus denoted Ul, U2, etc. The most abundant are those involved in pre-mRNA splicing Ut, U2, U4, US and U6. However, the list of snRNAs is growing, and the majority seem to be involved in determining the sites of methylation of prerRNA and are thus located in the nucleolus. The major nucleoplasmic snRNPs are formed by the individual snRNAs complexing with a common set of eight proteins, which are small and basic, and a variable number of snRNP-specific proteins. These core proteins, known as the **Sm proteins** (after an antibody which recognizes them), require the sequence 5'-$RA(U)_nGR$-3' to be present in a single-stranded region of the RNA. U6 does not have this sequence but it is usually base-paired to U4 which does. The snRNPs are formed as follows. They are synthesized in the nucleus by RNA Pol II and have a normal 5'-cap. They are exported to the cytoplasm where they associate with the common core proteins and with other specific proteins. Their 5'-cap gains two methyl groups and they are then imported back into the nucleus where they function in splicing.

4.1. Capping (5’ Capping):

Very soon after RNA Pol II starts making a transcript, and before the RNA chain is more than 20-30 nt long, the 5' -end is chemically modified by the addition of a 7-methylguanosine (m^7G) residue *(Fig.* 10). This 5'

modification is called a cap and occurs by addition of a GMP nucleotide to the new RNA transcript in the reverse orientation compared with the normal 3'-5' linkage, giving a 5'-5' Triphosphate Bridge. The reaction is carried out by an enzyme called mRNA guanyl transferase and there can be subsequent methylations of the sugars on the first and second transcribed nucleotides, particularly in vertebrates. The cap structure forms a barrier to 5'-exonucleases and thus stabilizes the transcript, but the cap is also important in other reactions undergone by pre-mRNA and mRNA, such as splicing, nuclear transport and translation.

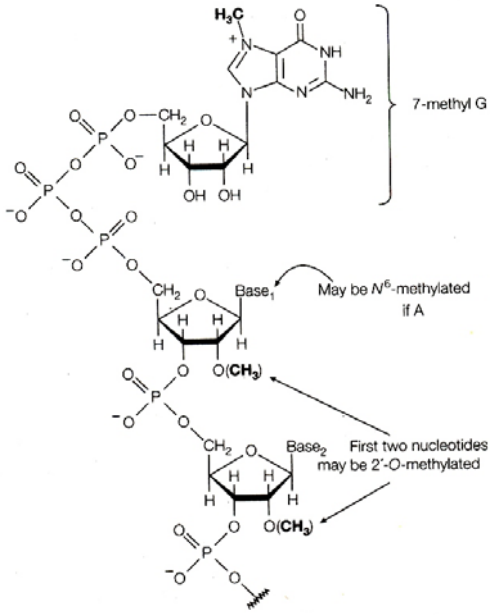


Fig. 10. The 5′ cap structure of eukaryotic mRNA.

4.2. Cleavage and polyadenylation at 3' end:

In most pre-mRNAs, the mature 3' -end of the molecule is generated by cleavage followed by the addition of a run, or tail, of A residues which is called the poly(A) tail. This feature has allowed the purification of mRNA molecules from the other types of cellular RNAs, permitting the construction of cDNA libraries from which specific genes have been isolated and their functions analyzed.

The cleavage and polyadenylation reaction requires that specific sequences be present in the DNA and its pre-mRNA transcript. These consist of a 5'-AAUAAA-3', the polyadenylation signal, followed by a 5'-YA-3', where Y = pyrimidine in the next 11-20 nt *(Fig 11).* Downstream, a GU-rich sequence is often present. Collectively, these sequence elements make up the requirements of a polyadenylation site.

A number of specific protein factors recognize these sequence elements and bind to the pre-mRNA. When the complex has assembled, cleavage takes place and then one of the factors, poly(A) polymerase (PAP), adds up to 250 A residues to the 3'-end of the cleaved pre-mRNA. The poly(A) tail on pre-mRNA is thought to help stabilize the molecule since a poly(A)-binding protein binds to it which should act to resist 3' -exonuclease action. In addition, the poly(A) tail may help in the translation of the mature mRNA in the cytoplasm. Histone pre-mRNAs do not get polyadenylated, but are cleaved at a special sequence to generate their mature 3' -ends.

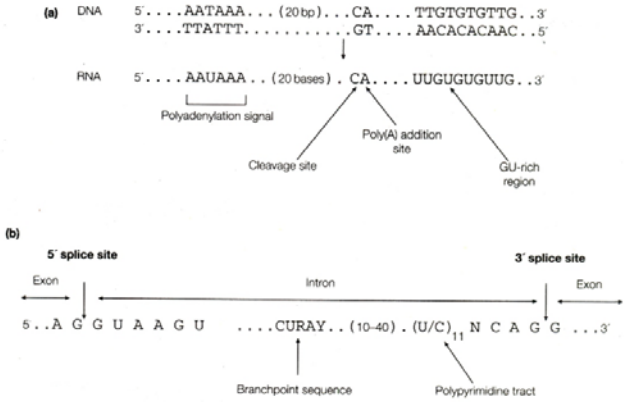


Fig. 11. Sequences of (a) a typical polyadenylation site and (b) the splice site consensus.

4.3. Splicing:

During the processing of pre-mRNA in eukaryotes, some sequences that are transcribed and which are upstream of the polyadenylation site are also eventually removed to create the mature mRNA. These sequences are cut out from central regions of the pre-mRNA and the outer portions joined. These intervening sequences, or introns, interrupt those sequences that will become adjacent regions in the mature mRNA, the exons which are usually the protein coding regions of the mRNA. The process of cutting the pre-mRNA to remove the introns and joining together of the exons is called splicing. Like the polyadenylation process, it takes place in the nucleus before the mature mRNA can be exported to the cytoplasm.

Splicing also requires a set of specific sequences to be present *(Fig 12).* The 5'-end of almost all introns has the sequence 5'-GU-3' and the 3'-end is usually 5'-AG-3'. The AG at the 3'-end is preceded by a pyrimidine-rich sequence called the **polypyrimidine tract**. About 10-40 residues upstream of the **polypyrimidine** tract is a sequence called the **branch point sequence** which is **5' -CURAY-3'** in vertebrates, where R = purine and Y = pyrimidine, but in yeast is the more specific sequence 5'-UACUAAC-3'.

Splicing has been shown to take place in a **two-step reaction** *(Fig 12.).* First, the bond in front of the G at the 5'-end of the intron at the so-called **5'-splice site** is attacked by the 2' -hydroxyl group of the A residue of the branchpoint sequence to create a tailed circular molecule called a **lariat** and free exon 1. In the second step, cleavage at the 3'-splice site occurs after the G of the AG, as he two exon sequences are joined together.

The intron is released in the lariat form and is eventually degraded.

The splicing process is catalyzed by the U1, U2, U4, U5 and U6 snRNPs, as well as other splicing factors. The RNA components of these snRNPs form base p)airs with the various conserved sequences at the 5'- and 3'-splice sites and he branchpoint *(Fig 12).* Early in splicing, the 5'-end of the U1 snRNP binds on the 5' -splice site and then U2 binds to the branchpoint. The tri-snRNP complex of U4, U5 and U6 can then bind, and in so doing the intron is looped our and the 5'- and 3'-exons are brought into close proximity. The snRNPs interact with one another forming a complex which folds the pre-mRNA into the correct conformation for splicing. This complex of snRNPs and pre-mRNA which forms to hold the upstream and downstream exons close together while looping out the intron is called a spliceosome. After the spliceosome forms, a rearrangement takes place before the .two-step splicing reaction can occur with release of the intron as a lariat.

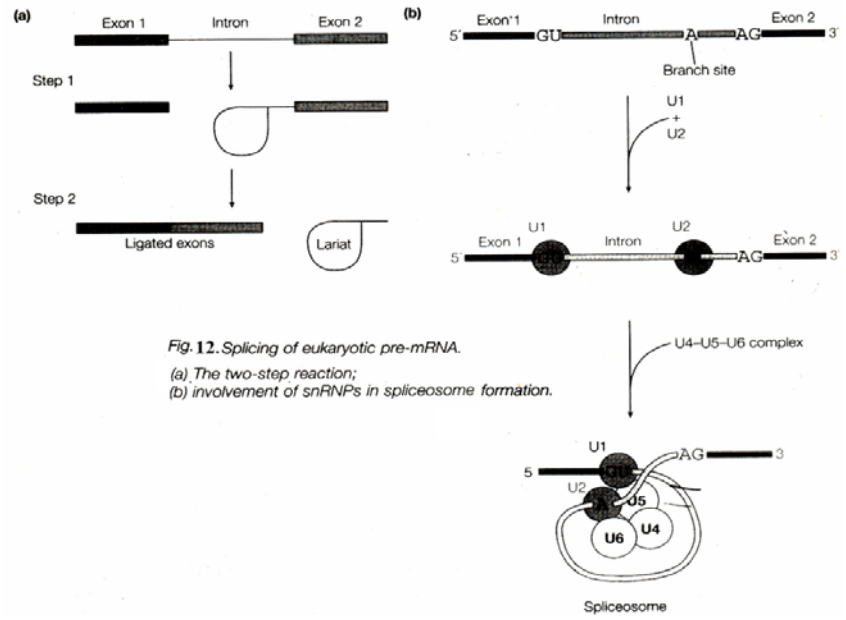


Fig. 12. Splicing of eukaryotic pre-mRNA.
(a) The two-step reaction;
(b) involvement of snRNPs in spliceosome formation.

Table : Properties of different classes of intron and summaries of their splicing mechanisms

Intron class	Characteristics of intron and splicing mechanism
Transesterification introns — Splice recognition sites within intron	
Nuclear pre-mRNA introns	Structurally diverse
	Initial donation of free hydroxyl group made by adenosine residue at internal branch site
	Lariat intermediate formed
	Large *trans*-acting splicing assembly required
Self-splicing introns class I	Conserved secondary structure
	Initial donation of free hydroxyl group made by guanosine nucleotide cofactor
	No lariat formed
	Autocatalytic splicing
Self-splicing introns class II	Conserved secondary structure
	Initial donation of free hydroxyl group made by adenosine residue at internal site
	Lariat intermediate formed
	Autocatalytic splicing
Self-splicing introns class III	Similar to group II introns but smaller (100–120 bp), containing a restricted number of domains
Twintrons	Multiple embedded self-splicing introns, often group II or mixed group II and group III. Usually spliced in a particular order
Nontransesterification introns — Splice recognition involves exon structure	
Nuclear tRNA introns	Splicing mechanism similar to tRNA maturation — involves cleavage followed by ligation
	No intermediate formed — intron excised as linear fragment
	Several processing enzymes required in *trans*

4.4. Pre-mRNA methylation:

The final modification or processing event that many pre-mRNAs undergo is specific methylation of certain bases. In vertebrates, the most common methylation event is on the N_6 position of A residues, particularly when these A residues occur in the sequence **5'-RRACX-3'**, where X is rarely G. Up to 0.1 % of pre-mRNA A residues are methylated, and the methylations seem to be largely conserved in the mature mRNA, though their function is unknown.

4.5. Alternative processing:

It has become clear that in many cases in eukaryotes a particular pre-mRNA species can give rise to more than one type of mRNA. This can occur when certain exons (alternative exons) are removed by splicing and so are not retained in the mature mRNA product. Additionally, if there are alternative possible poly(A) sites that can be used, different 3' -ends can be present in the mature mRNAs. Types of alternative RNA processing include alternative (or differential) splicing and alternative (or differential) poly(A) processing.

4.6. Alternative poly (A) sites: Some pre-mRNAs contain more than one set of the sequences required for cleavage and polyadenylation. The cell, or organism, has a choice of which one to use. It is possible that if the upstream site is used then sequences that control mRNA stability or location are removed in the portion that is cleaved off. Thus mature mRNAs with the same coding region, but differ-I\g stabilities or locations, could be produced from one gene. In some situations, both sites could be used in the same cell at a frequency that reflects their relative efficiencies (strengths) and the cell would contain both types of mRNA. The efficiency of a poly(A) site may reflect how well it matches the consensus sequences In other situations, one cell may exclusively use one poly(A) site, while a different cell uses another. The most likely explanation is that in one cell the stronger site is used by default, but in the other cell a factor is present that activates the weaker site so it is used exclusively, or that prevents the stronger site from being used. In some cases, the use of alternative poly(A) sites can cause different patterns of splicing to occur.

4.7. Alternative Splicing: Four common types of alternative splicing are summarized in *Fig 13*. In *Fig 13,* it is the choice of promoter that forces the pattern of splicing, as happens in the α-amylase and myosin light chain genes. The exon transcribed from the upstream promoter has the stronger 5' -splice site which out-competes the downstream one for use of the the first 3'-splice site. This happens in salivary gland for the α-amylase gene when specific transcription factors cause transcription from the upstream promoter. In the liver, the downstream promoter is used and the weaker (second) 5'-splice site is used by default. Alternative splicing caused by differential use of poly(A) sites is shown in *Fig 13.*. The stronger 3' -splice site is only present if the downstream poly(A) site is used and thus the penultimate 'exon' will be removed. When the upstream poly(A) site is used (such as in a different cell or at a different stage of development), splicing occurs by default using the weaker (upstream) 3'-splice site. In the case of immunoglobulins, use of a downstream poly(A) site includes exons encoding membrane-anchoring regions whereas when the upstream site is used these regions are not present and the secreted form of immunoglobulin is produced. In some situations, introns can be retained, as shown in *Fig.* If the intron contains a stop codon then a truncated protein will be produced on translation. This can give rise to an inactive protein, as in the case of the P element transposase in *Drosophila* somatic cells. In germ cells, a specific factor (or the lack of one present in somatic cells) causes the correct splicing of the intron and a longer mRNA is made which is translated into a functional enzyme in these cells only. The final type of alternative splicing *(Fig 13.)* illustrates that some exons can be retained or removed in different circumstances. A likely reason is the existence of a factor in one cell type that either promotes the use of a particular splice site or prevents the use of another. The rat troponin-T pre-mRNA can be differentially spliced in this way.

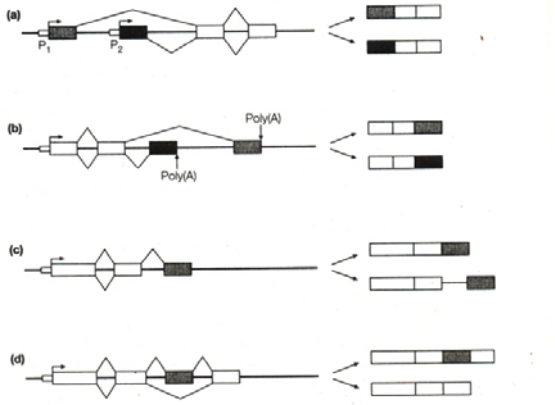


Fig. 13. Modes of alternative splicing. (a) Alternative selection of promoters P_1 or P_2; (b) alternative selection of cleavage/polyadenylation sites; (c) retention of an intron; (d) exon skipping. Empty boxes are exons, filled boxes are alternative exons and thin lines are introns.

4.8. RNA Editing: An unusual form of RNA processing in which the sequence of the primary transcript is altered is called RNA editing. Several examples exist, and they seem to be more common in non-vertebrates. In man, editing causes a single base change from C to U in the **apolipoprotein B** pre-mRNA, creating a stop codon in the mRNA in intestinal cells at position 6666 in the 14500 nt molecule. The unedited RNA in the liver makes apolipoprotein 8100, a 512 kDa protein, but in the intestine editing causes the truncated apolipoprotein 848 (241 kDa) to be made. Similarly, a single A to G change in the glutamate receptor pre-mRNA gives rise to an altered form of the receptor in neuronal cells. The RNA editing changes in the ciliated protozoan, *Leishmania,* are much more dramatic. When the cDNA for the mitochondrial cytochrome *b* gene was cloned, it had a coding region corresponding to the protein

sequence. However, although the gene was known to be encoded by the mitochondrial genome, no corresponding sequence was apparent. Eventually, some cDNA clones were obtained which had sequences corresponding to intermediates between the mature mRNA and the genomic sequence. It seems that the primary transcript is edited successively by introducing U residues at specific points. Many cycles of editing eventually produce the mature mRNA which can be translated. Short RNA molecules called guide RNAs seem to be involved. Their sequences are complementary to regions of the genomic DNA and the edited RNA. Several other types of RNA editing are known.

Table : Four RNA editing mechanisms

Editing process	Properties	Examples	Mechanism
Simple editing	Single residue conversions Posttranscriptional	C→U transition in the mammalian apolipoprotein mRNA	Modification by specific cytidine deaminase
		A→G transition in mRNA for mammalian glutamate receptor subunits and serotonin receptors. Editing also occurs in introns	Modification by dsRNA deaminase converts adenosine to *inosine* Three genes have been identified
		C→U transitions in plant organelles	Unknown
Insertional editing	Insertion of single nucleotides or small runs of nucleotides Cotranscriptional	G insertions during transcription of the paramyxovirus P gene	Transcriptional strand slipping
Pan-editing	Insertion/deletion of multiple uridine residues Posttranscriptional	U insertions/deletions in trypanosome kinetoplast mRNA	Editing sequence provided by external antisense **guide RNA (gRNA)** which pairs with **pre-edited mRNA** in a ribonucleoprotein particle, the **editosome**, and identifies positions to be edited as mismatches. gRNA has polyuridylate tail which supplies uridine residues for insertion. Editing has 3′→5′ polarity
	Insertion of multiple cytidine residues	C insertions in at least four *Phisarum polycephalum* mitochondrial mRNAs	Unknown
Polyadenylation editing	Addition of adenosine residues at end of transcript to complete stop codons	Polyadenylation of several vertebrate mitochondrial mRNAs	Pre-mRNA lacking a stop codon is polyadenylated, with the first one or two adenylate residues providing the missing information

B5. RNA TRANSPORT & POST-PROCESSING REGULATION

RNA export and sub-cellular localization:

In eukaryotes, mRNA is synthesized in the nucleus and must be transported to the cytoplasm for translation. It is thought that the restriction of RNA processing complexes to discrete foci in the nucleus plays a direct role in the subsequent export of RNA through nuclear pores. The exact mechanism of export is still not fully understood. It is known that there is selective transport of mRNPs, and that export is dependent upon ATP hydrolysis. It is also known that the **5'** cap plays a major role in nuclear export (uncapped transcripts such as rRNA and U6 snRNA are not exported), and that the presence of spliceosome components blocks export (thus preventing the translation of partially spliced transcripts). Some hnRNP proteins are removed from the transcript before export, and some dissociate following transport and return to the nucleus to be reused. Processed mRNA associates immediately with ribosomes as it leaves the nucleus.

RNA export from the nucleus represents a potential regulatory target in eukaryotes, but to date only viral RNA has been shown to be controlled in this manner. The best-characterized system is the ***HIV*** genome, where splicing and export are controlled by the Rev protein and cis-acting elements, Rev response elements, in the introns. Rev appears to facilitate the export of RNA with bound spliceosomes, and thus allows partially spliced ***HIV*** genomes to be exported. It is not clearly understood how Rev circumvents the normal nuclear inhibition of this process.

Once in the cytoplasm, mRNAs associated with ribosomes may diffuse freely, or may be targeted to a

particular region of the cell. Partially translated transcripts encoding secreted proteins are often transported to the membrane of the rough endoplasmic reticulum so that the polypeptide can be translocated into the lumen of this organelle (***signal peptide***). Whereas this relies upon a localization signal in the polypeptide, in other cases the signal is carried by the RNA itself (RNA targeting).

Certain transcripts become associated with the cytoskeleton, and are localized to specific regions of the cell. The latter is a common phenomenon during animal development as it provides a mechanism to localize positional signals as cytoplasmic determinants in the egg (e.g. bicoid and nanos mRNA in Drosophila development, ***vg-1*** mRNA in ***Xenopus*** development). In other cases, localization can be directly related to function

(e.g. the localization of α-actin and β-actin mRNAs during myoblast differentiation) or may be essential for cell survival (e.g. translation of myelin basic protein in the wrong location is lethal). In each case, the localization signal is found in the 3' UTR of the transcript and associates with one or more proteins required for its localization.

m-RNA turnover:

The abundance of a particular transcript is controlled both by its rate of synthesis and its stability, which reflects its rate of degradation. The stability of mRNA determines how quickly the steady-state levels of the mRNA change when the rate of transcription is altered, and thus how much is available for protein synthesis. Like other forms of gene regulation, RNA stability can be constitutive or regulated. Changes in the rate of mRNA degradation, which can be expressed as the mRNA half-life, can effect rapid and transient alterations to the abundance of a particular mRNA without any change in transcriptional activity.

m-RNA turnover and retro-regulation* in *bacteria: Bacterial mRNAs have short half-lives, in the order of several minutes for the most stable transcripts, which allows the rate of protein synthesis to be altered rapidly in response to the environment by regulating the rate of transcription. mRNA degradation in prokaryotes is mediated by RNA endonucleases (also called RNA restriction enzymes) and *3'→5'* exonucleases. The secondary structure of mRNA is important in the determination of stability, with the most stable transcripts possessing multiple hairpins and stem-loop structures in the 3' untranslated region which may protect the transcript from exonuclease activity. Transcripts which contain endonuclease target sites are particularly unstable. The specific enzymes which degrade RNA in bacteria are not well characterized, although mutations which disrupt the *E.* coli ribonuclease E protein induce a 2-3-fold increase in RNA stability.

The close association of transcription, protein synthesis and degradation in bacteria permits an unusual form of gene regulation, termed retro-regulation, where RNA degradation is regulated at the level of transcriptional termination. Gene expression depends upon whether or not a cis-acting element located downstream of the gene is transcribed. If it is, the nascent RNA adopts a structure favoring rapid degradation and translation is prevented. If transcription terminates prior to this site, the RNA is relatively stable and protein synthesis proceeds. Retro-regulation is used, for example, by bacteriophage λ to control the expression of its integrase gene.

m-RNA turnover in eukaryotes: Eukaryotic mRNAs are generally much more stable than bacterial transcripts. The half-life of yeast mRNA ranges from -5 to -45 minutes, and metazoan mRNA is even more stable, with an average half-life of about 10 hours, reflecting the relatively constant environment of cells in multicellular organisms. The polyadenylate tail present on most mammalian mRNAs appears to confer stability by binding the PABP, which maintains tail length. Deadenylation or depletion of PABP results in rapid mRNA degradation in mammalian cells, but in yeast cells, the presence of PABP appears to be a signal for degradation, so its precise role is unclear. In eukaryotes, exoribonuclease degradation of the polyadenylate tail is the first stage in mRNA degradation.

Histone mRNAs, which lack polyadenylate tails, are degraded by a specific exonuclease. There is also evidence for endoribonucleases active in eukaryotic cells. Some eukaryotic regulatory proteins are required in transient bursts, and their mRNAs are consequently unstable like bacterial transcripts. These include the transcripts of many immediate early genes, the genes induced by signal transduction cascades and required to produce a short-lived regulatory responses (e.g. c-fos, c-jun; Oncogenes and Cancer, Signal Transduction).

Many unstable eukaryotic mRNAs contain specific instability elements, often AU-rich elements (AUREs, ARES) such as multiple copies of the sequence AUUUA, generally located in the 3' UTR (e.g. in interleukin 1, interferon β and c-fos mRNAs), although in several cases, within the coding region (e.g. in β-tubulin and c-myc mRNAs). Although the mechanism by which instability is conferred is not understood, AUREs have the ability to form stem-loop structures, suggesting that factors which influence RNA folding may regulate RNA stability. Instability elements appear to be relatively independent, as they can confer instability on heterologous mRNA when inserted into the 3' untranslated region. However, there may be some dependence on secondary structure and/or context as similar elements have been identified in a number of stable transcripts such as neuronspecific enolase mRNA.

The stability of a few eukaryotic mRNAs can be regulated by trans-acting factors which bind to the instability elements. This occurs in the transferrin

receptor mRNA which, in the presence of excess intracellular iron levels, is degraded, presumably by the same process which controls degradation of the constitutively unstable transcripts described above. When iron levels fall, however, a protein factor binds to an iron response element (IRE) in the transcript and prevents degradation.

IRES are 30 bp motifs which form stem-loop structures with instability elements in the stem. The presence of the IRE-binding protein is thought to block access to the instability elements. In other transcripts, the same IRE is used to regulate translation.

C1. RIBOSOMES (RNPs) AND THEIR STUDY

The RNA molecules in cells usually exist complexed with proteins, specific proteins attaching to specific RNAs. These RNA-protein complexes are called **ribonucleoproteins (RNPs)**. **Ribosomes** are the largest and most complex RNPs and are formed by the rRNA molecules complexing with specific ribosomal proteins during processing. Several methods are used to study RNPs, including dissociation, where the RNP is purified and separated into its RNA and protein components which are then characterized. **Re-assembly** is used to discover the order in which the components fit together and, if the components can be modified, it is possible to gain clues as to their individual functions. Electron microscopy can allow direct visualization if the RNPs are large enough, otherwise it can roughly indicate overall shape. **Antibodies to RNPs** or their individual components can be used for purification, inhibition of function and, in combination with electron microscopy; they can show the crude positions of the components in the overall structure. **RNA binding experiments** can show whether a particular protein binds to an RNA, and subsequent treatment of the RNA-protein complex with RNase (**RNase protection experiment**) can show which parts of the RNA are protected by bound protein (i.e. the site of binding). **Cross-linking experiments** using UV light with or without chemical cross-linking agents can show which parts of the RNA and protein molecules are in close contact in the complex. Physical methods such as **neutron** and **X-ray diffraction** can ultimately give the complete 3-D structure. Collectively, these methods have provided much information on the structure of the RNPs.

Prokaryotic ribosomes: The importance of ribosomes to a cell is well illustrated by the fact that in E. *coli* ribosomes account for 25% of the dry weight (10% of total protein and 80% of total RNA). *Figure 14* shows the components present in the E. *coli* ribosome. The 70S ribosome of molecular mass 2.75 X 10^6 Da is made up of a large subunit of 50S and a small subunit of 30S. The latter is composed of one copy of the 16 S rRNA molecule and 21 different proteins denoted S_1 to S_{21}, The large subunit contains one 23 S and one 5 S rRNA molecule and 31 different proteins. These were named L_1 to L_{34} after fractionation on two-dimensional gels. However, the L_{26} spot was later found to be S_{20}, L_7 is the acetylated version of L_{12} and L_8 is a complex of L_{10} and L_7, hence there are only 31 different large subunit proteins. The sizes of these ribosomal proteins vary widely, from L_{34} which is only 46 amino acids to S_1 which is 557. Mostly, these relatively small proteins are basic, which might be expected since they bind to RNA. It is possible to re-assemble functional E. *coli* ribosomes from the RNA and protein components, and there is a defined pathway of assembly. The various methods of studying RNPs have led to the structures shown in *figure 15* for the *E. coli* ribosomal subunits.

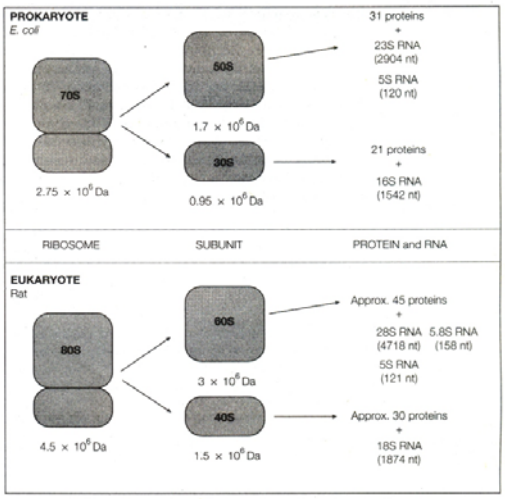


Fig. **14.** *Composition of typical prokaryotic and eukaryotic ribosomes.*

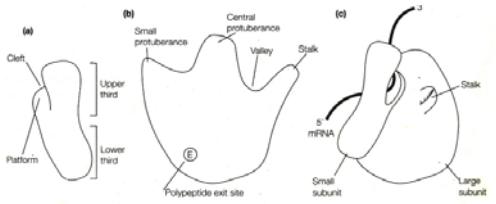


Fig. **15.** *Features of the E. coli ribosome.*
(a) The 30S subunit; (b) the 50S subunit; (c) the complete 70S ribosome.

Eukaryotic ribosomes: The corresponding sizes of, and components in the 80 S eukaryotic (rat) ribosome are shown in *Fig 14.* In the large 60 S subunit, which contains about 45 proteins and one 5 S rRNA molecule, the 5.8 S rRNA and the 28 S rRNA molecules together are the equivalent of the prokaryotic 235 molecule. The small 40S subunit contains the 18 S rRNA and about 30 different proteins. Although all the rRNAs are larger in eukaryotes, there is a considerable degree of conser-vation of secondary structure in each of the corresponding molecules. Due to their greater complexity, the eukaryotic ribosomal subunits have not yet been re-assembled into functional complexes and their structure is less well understood. The ribosomes in a typical eukaryotic cell can collectively make about a million peptide bonds per second.

C2. GENETIC CODE

The genetic code is the correspondence between the sequence of the four bases in nucleic acids and the sequence of the 20 amino acids in proteins. It has been shown that the code is a triplet code, where three nucleotides encode one amino acid, and this agrees with mathematical argument as being the minimum necessary [(4^2 = 16) < 20 < (4^3 = 64)]. However, since there are only 20 amino acids to be specified and potentially 64 different triplets, most amino acids are specified by more than one triplet and the genetic code is said to be **degenerate**, or to have **redundancy**. From a fixed start point, each group of three bases in the coding region of the mRNA represents a codon which is recognized by a complementary triplet, or **anticodon**, on the end of a particular tRNA molecule. The triplets are read in non-overlapping groups and there is no punctuation between the codons to separate or delineate them. They are simply decoded as adjacent triplets once the process of decoding has begun at the correct start point. As more gene and protein sequence information has been obtained, it has become clear that the genetic code is very nearly, but not quite, universal. This supports the hypothes.is that all life has evolved from a single common origin.

2.1 Deciphering: In the 1960s, Nirenberg developed a cell-free protein synthesizing system from E. *coli.* Essentially, it was a centrifuged cell lysate which was DNase treated to prevent new transcription and which would carry out limited protein synthesis if natural or synthetic mRNA was added. To determine which amino acids were being polymerized into polypeptides, it was necessary to carry out 20 reactions in parallel. Each reaction had 19 nonradioactive amino acids and one amino acid labeled with radioactivity. The enzyme poly. **nucleotide phosphorylase** was used to make synthetic mRNAs that were composed of only one nucleotide, that is poly(U), poly(C), poly(A) and poly(G). If protein synthesis took place after adding one of these homopolymeric synthetic mRNAs, then in one of the 20 reaction tubes radioactivity would be incorporated into polypeptide. In this way, it was found that poly(U) caused the synthesis of polyphenylalanine, poly(C) coded for polyproline and poly(A) for polylysine. Poly(G) did not work because it formed a complex secondary structure.

If polynucleotide phosphorylase is used to polymerize a mixture of two nucleotides, say U and G, at unequal ratios such as 0.76: 0.24, then the triplet GGG is the rarest and UUU will be most common. Triplets with two Us and one G will be the next most frequent. By using these **random co-polymers** as synthetic mRNAs in the cell-free system and determining the frequency of incorporation of particular amino acids, it was possible to determine the composition of the codon for many amino acids. The precise sequence of the triplet codon can only be worked out if additional information is available.

Towards the end of the 1960s, it was found that synthetic trinucleotides could attach to the ribosome and bind their corresponding aminoacyl-tRNAs from a mixture. Upon filtering through a membrane, only the complex of ribosome, synthetic triplet and aminoacyl-tRNA was retained on the membrane. If the mixture of aminoacyl-tRNAs was made up 20 times, but each time with a different radioactive amino acid, then in this experiment specific triplets could be assigned unambiguously to specific amino acids. A total of 61 codons were shown to code for amino acids and there were three stop codons *(Table 1*)

Table 1. The universal genetic code

First position (5′ end)	Second position U	Second position C	Second position A	Second position G	Third position (3′ end)
U	Phe UUU	Ser UCU	Tyr UAU	Cys UGU	U
	Phe UUC	Ser UCC	Tyr UAC	Cys UGC	C
	Leu UUA	Ser UCA	**Stop** UAA	**Stop** UGA	A
	Leu UUG	Ser UCG	**Stop** UAG	Trp UGG	G
C	Leu CUU	Pro CCU	His CAU	Arg CGU	U
	Leu CUC	Pro CCC	His CAC	Arg CGC	C
	Leu CUA	Pro CCA	Gln CAA	Arg CGA	A
	Leu CUG	Pro CCG	Gln CAG	Arg CGG	G
A	Ile AUU	Thr ACU	Asn AAU	Ser AGU	U
	Ile AUC	Thr ACC	Asn AAC	Ser AGC	C
	Ile AUA	Thr ACA	Lys AAA	Arg AGA	A
	Met AUG	Thr ACG	Lys AAG	Arg AGG	G
G	Val GUU	Ala GCU	Asp GAU	Gly GGU	U
	Val GUC	Ala GCC	Asp GAC	Gly GGC	C
	Val GUA	Ala GCA	Glu GAA	Gly GGA	A
	Val GUG	Ala GCG	Glu GAG	Gly GGG	G

2.2 Features: The genetic code is degenerate (or it shows redundancy). This is because 18 out of 20 amino acids have more than one codon to specify them, called **synonymous codons**. Only methionine and tryptophan have single codons. The synonymous codons are not positioned randomly, but are grouped in the table. Generally they differ only in their third position. In all cases, if the third position is a pyrimidine, then the codons specify the same amino acid (are synonymous). In most cases, if the third position is a purine the codons are also synonymous. If the second position is a pyrimidine then generally the amino acid specified is hydrophilic. If the second position is a purine then generally the amino acid specified is polar.

2.3 Effect of mutation: It is generally considered that the genetic code evolved in such a way as to minimize the effect of mutations. The most common type of mutation is a transition, where either a purine is mutated to the other purine or a pyrimidine is changed to the other pyrimidine. Transversions are where a pyrimidine changes to a purine or vice versa. In the third position, transitions usually have no effect, but can cause changes between Met and Ile, or Trp and stop. Just over half of transversions in the third position have no effect and the remainder usually results in a similar type of amino acid being specified, for example Asp or Glu. In the second position, transitions will usually result in a similar chemical type of amino acid being used, but transversions will change the type of amino acid. In the first position, mutations (both transition and transversions) usually specify a similar type of amino acid, and in a few cases it is the same amino acid.

2.4 Universality: For a long time after the genetic code was deciphered, it was thought to be universal, that is the same in all organisms. However, since 1980, it has been discovered that mitochondria, which have their own small genomes, utilize a genetic code that differs slightly from the standard or 'universal' code. Indeed, it is now known that some other unicellular organisms also have a variant genetic code. *Table 2* shows the variations in the genetic code.

2.5 ORFs: Inspection of DNA sequences, such as those obtained by genome sequencing projects, by eye or by computer will identify continuous groups of adjacent codons that start with ATG and end with TGA, TAA or TAG. These are referred to as open reading frames, or ORFs, when there is no known protein product. When a particular ORF is known to encode a certain protein, the ORF is usually referred to as a coding region. Hence, an ORF is a suspected coding region.

Table 2 Modifications of the genetic code

Codon	Usual meaning	Alterative	Organelle or organism
AGA	Arg	Stop, Ser	Some animal mitochondria
AGG			
AUA	Ile	Met	Mitochondria
CGG	Arg	Trp	Plant mitochondria
CUN	Leu	Thr	Yeast mitochondria
AUU	Lie	Start	Some prokaryotes
GUG	Val		
UUG	Leu		
UAA	Stop	Glu	Some protozoans
UAG			
UGA	Stop	Trp	Mitochondria, mycoplasma

2.6 Overlapping genes: Although it is generally true that one gene encodes one polypeptide, and the evolutionary constraints on having more than one protein encoded in a given region of sequence are great, there are now known to be several examples of overlapping coding regions (overlapping genes). Generally these occur where the genome size is small and there is a need for greater information storage density. For example, the phage ϕX174 makes 11 proteins of combined molecular mass 262 kDa from a 5386 bp genome. Without overlapping genes, this genome could encode at most 200 kDa of protein. Three proteins are encoded within the coding regions for longer proteins. In prokaryotes, the ribosomes simply have to find the second start codon to be able to translate the overlapping gene and they may achieve this without detaching from the template. Eukaryotes have a different way of initiating protein synthesis and tend to make use of alternative RNA processing to generate variant proteins from one gene.

C3. AMINOACYLATION OF t-RNA, t-RNA-IDENTITY, AMINOACYL t-RNA SYNTHETASE, TRANSLATIONAL PROOF-READING

3.1 t-RNA Structure and Functions:

t-RNA primary Structure: tRNAs are the adaptor molecules that deliver amino acids to the ribosome and decode the information in mRNA. Their primary structure (i.e. the linear sequence of nucleotides) is 60-95 nt long, but most commonly 76. They have many modified bases sometimes accounting for 20% of the total bases in anyone tRNA molecule. Indeed, over 50 different types of modified base have been observed in the several hundred tRNA molecules characterized to date, and all of them are created post-transcriptionally. Seven of the most common types are shown in *Fig 16.* as nucleosides. Four of these, ribothymidine (T), which contains the base thymine not usually found in RNA, pseudouridine (ψ), dihydrouridine (D) and inosine (I), are very common in tRNA, all but the last being present in nearly all tRNA molecules in similar positions in the sequences. The letters D and T are used to name secondary structural features.

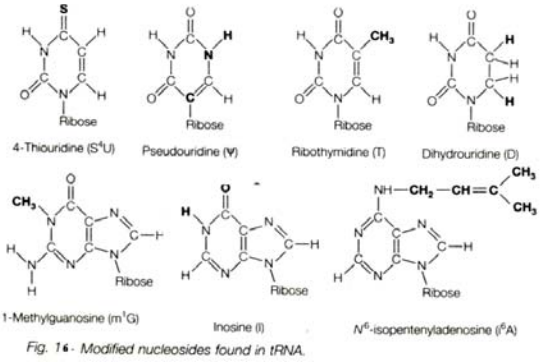


Fig. 16 - Modified nucleosides found in tRNA.

In the tRNA primary structure, there are 15 invariant nucleotides and eight which are either purines (R) or pyrimidines (Y). Using the standard numbering convention where position 1 is the 5' -end and 76 is the 3'-end, these are:

14,15, 18, 19, 21,	**D-Loop**
32,33,37,	**Anticodon**
53,54,55,56,57,58,60,61,	**T-loop**
74,75,76	**acceptor loop**

The positions of invariant and semi-variant nucleotides play a role in either the secondary or tertiary structure.

t-RNA secondary structure: All tRNAs have a common secondary structure (i.e. base pairing of different regions to form stems and loops), the cloverleaf structure shown in *Fig 17 (a).* This structure has a 5' -phosphate formed by RNase P cleavage, not the usual 5'-triphosphate. It has a 7 bp stem formed by base pairing between the 5'and 3'-ends of the tRNA; however, the invariant residues 74-76 (i.e. the terminal 5'-CCA-3') which are added during processing in eukaryotes) are not included in this base pairing region. This stern is called the amino acid acceptor stem. Working 5' to 3' (anticlockwise), the next secondary structural feature is called the D-arm which is composed of a 3 or 4 bp stem and a loop called the D-loop (DHU-loop) usually containing the modified base dihydrouracil. The next structural feature consists of a 5 bp stem and a seven residue loop in which there are three adjacent nucleotides called the anticodon which are complementary to the codon sequence (a triplet in the mRNA) that the tRNA recognizes. The presence of inosine in the anticodon gives a tRNA the ability to base-pair to more than one codon sequence. Next there is a variable arm which can have between three and 21 residues and may form a stem of up to 7 bp. The other positions of length variation in tRNAs are in the D-loop shown as dashed lines in *Fig 17a.* The final major feature of secondary structure is the T-arm or TψC-arm which is composed of a 5 bp stem ending in a loop containing the invariant residues GTψC. Note that the majority of the invariant residues in tRNA molecules are in the loops and do not play a major role in forming the secondary structure. Several of them help to form the tertiary structure.

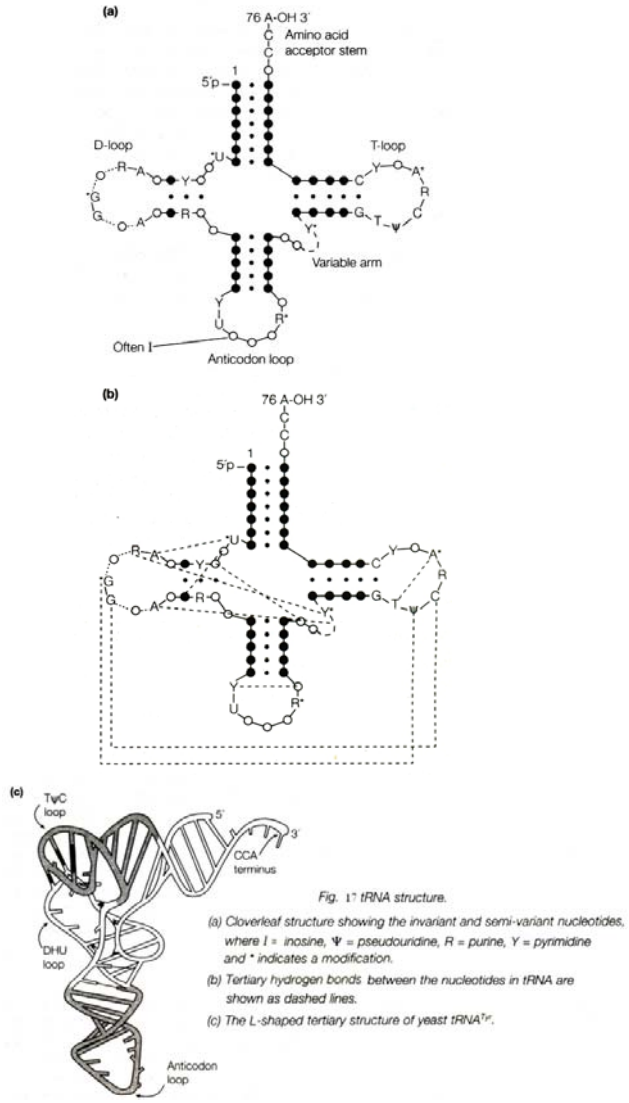


Fig. 17 tRNA structure.
(a) Cloverleaf structure showing the invariant and semi-variant nucleotides, where I = inosine, Ψ = pseudouridine, R = purine, Y = pyrimidine and • indicates a modification.
(b) Tertiary hydrogen bonds between the nucleotides in tRNA are shown as dashed lines.
(c) The L-shaped tertiary structure of yeast tRNA^{Phe}.

t-RNA tertiary structure: There are nine hydrogen bonds (tertiary hydrogen bonds) that help form the 3-D structure of tRNA molecules. They mainly involve base pairing between several invariant bases and are shown in *Fig. 17 b.* Base pairing between residues in the D- and T -arms fold the tRNA molecule over into an L-shape, with the anticodon at one end and the amino acid acceptor site at the other. The tRNA tertiary structure is strengthened by base stacking interactions *(Fig. 17 c)*.

t-RNA function: tRNAs are joined to amino acids to become aminoacyl-tRNAs (charged tRNAs) in a reaction called aminoacylation. It is these charged tRNAs that are the adaptor molecules in protein synthesis. Special enzymes called aminoacyl-tRNA synthetases carry out the joining reaction which is extremely specific (i.e. a specific amino acid is joined to a specific tRNA). These pairs of specific amino acids and tRNAs, or tRNAs and aminoacyl-tRNA synthetases are called cognate pairs, and the nomenclature used is shown in *Table 1.*

Table: Nomenclature of tRNA-synthetases and charged tRNAs

Serine	$tRNA^{Ser}$	Seryl- tRNA synthetase	Seryl-$tRNA^{Ser}$
Leucine	$tRNA^{Leu}$	Leucyl- tRNA synthetase	Leucyl-$tRNA^{Leu}$
	$tRNA^{Leu}_{UUA}$		Leucyl-$tRNA^{Leu}_{UUA}$

3.2 Amino-acylation of t-RNAs:

The general amino acylation reaction is shown in *Fig 18.* It is a two-step reaction driven by ATP. In the first step, AMP is linked to the carboxyl group of the amino acid giving a high-energy intermediate called an aminoacyl adenylate. The hydrolysis of the pyrophosphate released (to two molecules of inorganic phosphate) drives the reaction forward. In the second step, the aminoacyl adenylate reacts with the appropriate uncharged tRNA to give the aminoacyl-tRNA and AMP. Some synthetases join the amino acid to the 2'-hydroxyl of the ribose and some to the 3'-hydroxyl, but once joined the two species can interconvert. The formation of an aminoacyl-tRNA helps to drive protein synthesis as the aminoacyl-tRNA bond is of a higher energy than a peptide bond and thus peptide bond formation is a favorable reaction once this energy-consuming step has been performed.

3.3. Amino-acyl-tRNA synthetases:

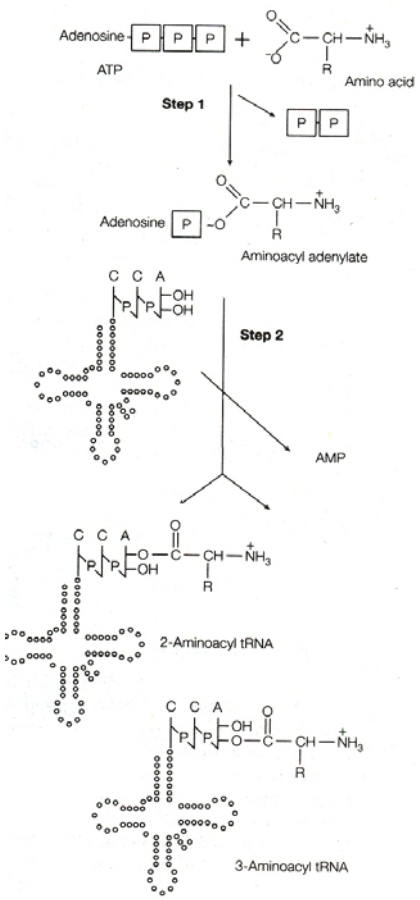


Fig 18 Formation of aminoacyl-tRNA.

Despite the fact that they all carry out the same reaction of joining an amino acid to a tRNA, the various synthetase enzymes can be quite different. They fall into one of four classes of subunit structure, being either, α2, α4 or α2-β2. The polypeptide chains range from 334 to over 1000 amino acids in length, and these enzymes contact the tRNA on the underside (in the angle) of the L-shape. They have a separate amino acid-binding site. The synthetases have to be able to distinguish between about 40 similarly shaped, but different, tRNA molecules in cells, and they use particular parts of the tRNA molecules, called identity elements, to be able to do this. These are not always the anticodon sequence (which does differ between tRNA molecules). They often include base pairs in the acceptor stem, and if these are swapped between tRNAs then the synthetase enzymes can be tricked into adding the amino acid to the wrong tRNA. For example, if the G3:U70 identity element of $tRNA^{Ala}$ is used to replace the 3:70 base pair of either $tRNA^{cys}$ or $tRNA^{Phe}$, then these modified tRNAs are recognized by alanyl-tRNA synthetase and charged with alanine.

3.4 Proof-reading: Some synthetase enzymes that have to distinguish between two chemically similar amino acids can carry out a proofreading step. If they accidentally carry out step 1 of the amino acylation reaction with the wrong amino acid, then they will not carry out step 2. Instead they will hydrolyze the amino acid adenylate. This proofreading ability is only necessary when a single recognition step is not sufficiently discriminating. Discrimination between the amino acids Phe and Tyr can be achieved in one step because of the -OH group difference on the benzene ring, so in this case there is no need for proofreading.

3.5 Aspects of Protein Synthesis:

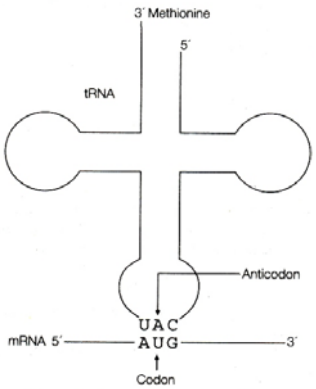


Fig. 19 Codon–anticodon interaction.

Codon-Anticodon Interaction: The anticodon at one end of the tRNA interacts with a complementary triplet of bases on the mRNA, the codon, when both are brought together in the cleft of 'the ribosome. The interaction is antiparallel in nature *(Fig.19*). Some highly purified tRNA molecules were found to interact with more than one codon, and this ability correlated with the presence of modified nucleosides in the 5'-anticodon position, particularly inosine. Inosine is formed by post-

transcriptional processing of adenosine if it occurs at this position. This is carried out by anticodon deaminase which converts the 6-amino group to a keto group.

Wobble:

The wobble hypothesis was suggested by Crick to explain the redundancy of the genetic code. He realized, by model building, that the 5' -anticodon base was able to undergo more movement than the other two bases and could thus form non standard base pairs as long as the distances between the ribose units were close to normal. His specific predictions are shown in *Table* along with actual observations.

No purine-purine or pyrimidine-pyrimidine base pairs are allowed as the ribose distances would be incorrect. No single tRNA could recognize more than three codons. Hence, at least 32 tRNAs would be needed to decode the 61 codons, excluding stop codons. tRNAs can recognize either one, two or three codons, depending on their wobble base (the 5' -anticodon base). If it is C it will recognize only the codon ending in G. If it is G, it will recognize the two codons ending in U or C. If U, which is subsequently modified, it will pair with either A or G. The wobble nucleoside is never A, as this is converted to inosine which then pairs with A, C or U.

Table 3. Original wobble predictions

5'anticodon base	Predicted 3' codon	Observation
A	**U**	A converted to I by anticodon deaminase
C	**G**	No wobble, normal base pairing
G	**C and U**	G, and modified G, can pair with C and U
U	**A**	not found as 5'-anticodon base
I	**A, C, U**	Wobble as predicted. Inosine (I) can recognize 3' -A, -C or -U

Ribosome Binding Site:

In prokaryotic mRNAs there is a conserved sequence 8-13 nt upstream of the first codon to be translated (the initiation codon). It was discovered by Shine and Dalgarno and is a purine-rich sequence usually containing all or part of the sequence 5' *-AGGAGGU-3'.* Experiments have shown that these sequences can base-pair with the 3'-end of the 16S rRNA in the small subunit of the ribosome *(5'-ACCUCCU-3').* It is called the ribosome binding site, or **Shine-Dalgarno** sequence. It is thought to position the ribosome correctly with respect to the initiation codon.

Polysomes:

When a ribosome has begun translating an mRNA molecule, and has moved about 70-80 nt from the initiation codon, a second ribosome can assemble at the ribosome-binding site and start to translate the mRNA. When this second ribosome has moved along, a third can begin and so on. Multiple ribosomes on a single mRNA are called polysomes (short for polyribosomes) and there can be as many as 50 on some mRNAs, although they cannot be positioned closer than about 80 nt.

Initiator tRNA:

It has been shown that the first amino add incorporated into a protein chain is methionine in both prokaryotes and eukaryotes, though in the former the Met has been modified to **N-formyl methionine**. In both types of organisms, the *AUG* initiation codon is recognized by a special initiator tRNA. The initiator tRNA differs from the one that pairs with *AUG* codons in the rest of the coding region. In the prokaryote E. *coli,* there are subtle differences between these two tRNAs *(Fig 20*).

The initiator tRNA allows more flexibility in base pairing (wobble) because it lacks the alkylated A in the anticodon loop and hence it can recognize both AUG and GUG as initiation codons, the latter occurring occasionally in prokaryotic mRNAs. The non-initiator tRNA is less flexible and can only pair with AUG codons. Both tRNAs are charged with Met by the same **methionyl-tRNA synthetase** to give the methionyl-tRNA, but only the initiator methionyl-tRNA is modified by the enzyme transformylase to give N-formylmethionyl-tRNAfMet. The N-formyl group resembles a peptide bond and may help this initiator tRNA to enter the P-site of the ribosome whereas all other tRNAs enter the A-site.

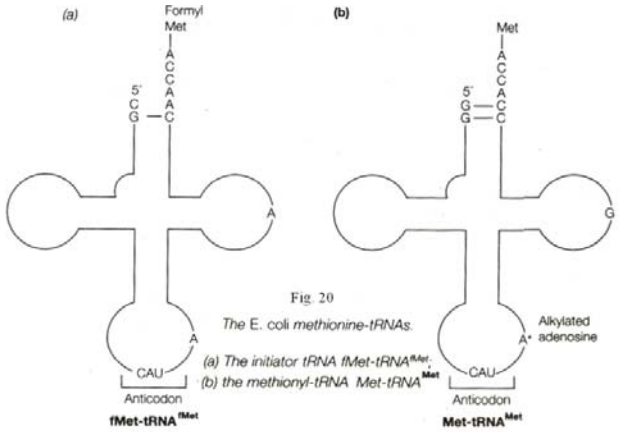


Fig. 20
The E. coli methionine-tRNAs.
(a) The initiator tRNA fMet-tRNA fMet.
(b) the methionyl-tRNA Met-tRNA Met

C4. MECHANISM OF PROTEIN SYNTHESIS IN PROKARYOTES

The actual mechanism of protein synthesis can be divided into three stages:

- **initiation** - the assembly of a ribosome on an mRNA molecule;
- **elongation** - repeated cycles of amino acid addition;
- **termination** - the release of the new protein chain.

These are illustrated in *Figs* and involve the activities of a number of factors. In prokaryotes, the factors are abbreviated as IF or EF for initiation and elongation factors respectively, whereas in eukaryotes they are called eIF and eEF. There are distinct differences of detail between the mechanism in prokaryotes and eukaryotes, and most of these occur in the initiation stage. For this reason, this topic will describe the mechanism in prokaryotes and the following topic will describe the differences in detail that occur in eukaryotes.

Initiation: The purpose of the initiation step is to assemble a complete ribosome on to an mRNA molecule at the correct start point, the initiation codon. The components involved are the large and small ri6osome subunits, the mRNA, the initiator tRNA in its charged form, three initiation factors and GTP. The initiation factors IF_1, IF_2 and IF_3 are all just over one-tenth as abundant as ribosomes, and have masses of 9, 120 and 22 kDa respectively. Only IF2 binds GTP. Although the finer details have yet to be worked out, the overall sequence of events *(Fig.* 1) is as follows:

- IF_1 and IF_3 bind to a free 30 S subunit. This helps to prevent a large subunit binding to it without an mRNA molecule and forming an inactive ribosome.
- IF2 complexed with GTP then binds to the small subunit. It will assist the charged initiator tRNA to bind.
- The 30S subunit attaches to an mRNA molecule making use of the ribosome binding site (RBS) on the mRNA.
- The initiator tRNA can then bind to the complex by base pairing of its anticodon with the AUG codon on the mRNA. At this point, IF_3 can be released, as its roles in keeping the subunits apart and helping the mRNA to bind are complete. This complex is called the 305 initiation complex.
- The 50s subunit can now bind, which displaces IF! and IF2' and the GTP is hydrolyzed in this energy-consuming step. The complex formed at the end of the initiation phase is called the 705 initiation complex.

As shown in *Figs* 1-3, the assembled ribosome has two tRNA-binding sites. These are called the A- and P-sites, for aminoacyl and peptidyl sites. The Asite is where incoming aminoacyl-tRNA molecules bind, and the P-site is where the growing polypeptide chain is usually found. These sites are in the cleft of the small subunit and contain adjacent codons that are being translated. One major outcome of initiation is the placement of the initiator tRNA in the P-site. It is the only tRNA that does this, as all others must enter the A-site.

(i) Aminoacyl-tRNA delivery: EF-Tu is required to deliver the aminoacyltRNA to the A-site and energy is consumed in this step by the hydrolysis of GTP. The released EF-Tu'GDP complex is regenerated with the help of EF- Ts. In the EF- Tu-EF- Ts exchange cycle, EF- Ts displaces the GDP and subsequently is displaced itself by GTP. The resultant EF- Tu .GTP complex is now able to bind another aminoacyl-tRNA and deliver it to the ribosome. All aminoacyl-tRNAs can form this complex with EF- Tu, except the initiator tRNA.

(ii) Peptide bond formation: After aminoacyl-tRNA delivery, the A- and P-sites are both occupied and the two amino acids that are to be joined are in close proximity. The peptidyl transferase activity of the 50S subunit can now form a peptide bond between these two amino acids without the input of any more energy, since energy in the form of A TP was used to charge the tRNA.

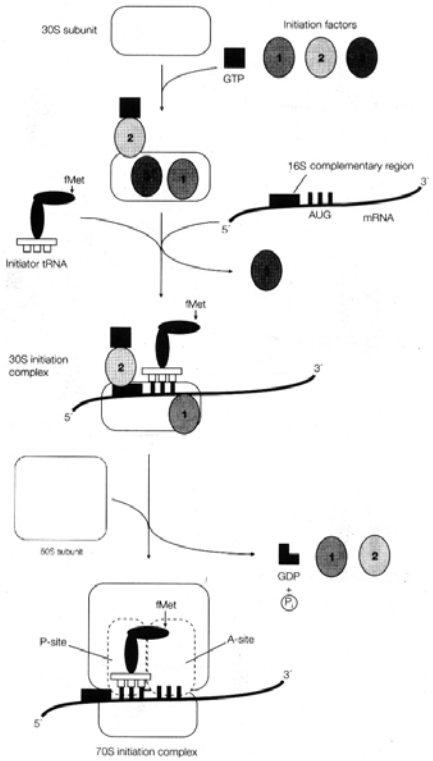


Fig. 1. Initiation of protein synthesis in the prokaryote E. coli.

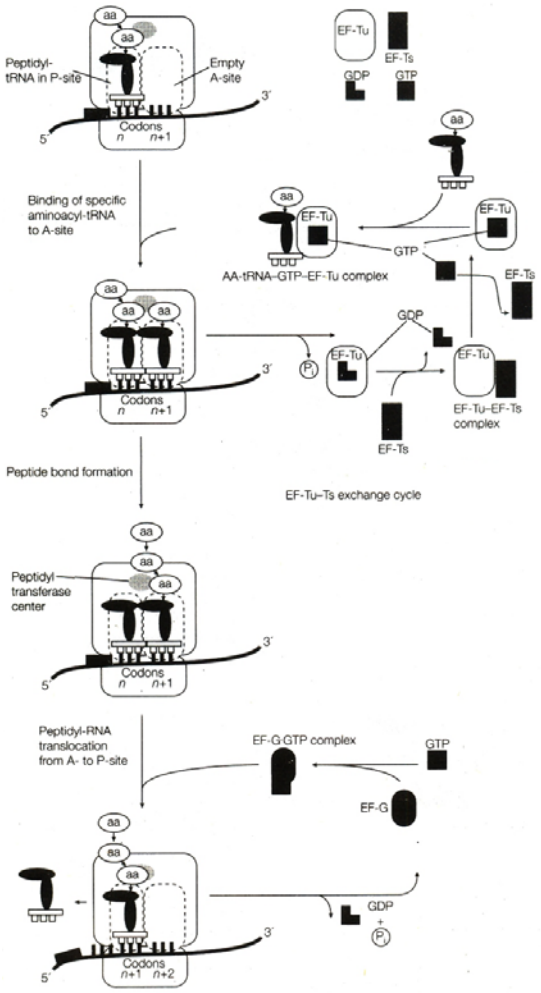


Fig. 2. Elongation stage of protein synthesis.

(iii) Translocation: A complex of EF-G (translocase) and GTP binds to the ribosome and, in an energy-consuming step, the discharged tRNA is ejected from the P-site, the peptidyl-tRNA is moved from the A-site to the P-site and the mRNA moves by one codon relative to the ribosome. GDP and EF-G are released, the latter being re-usable. A new codon is now present in the vacant A-site. Recent evidence suggests that in prokaryotes the discharged tRNA is first moved to an E-site (exit site) and is ejected when the next aminoacyl-tRNA binds. In this way the ribosome maintains contact with the mRNA via 6 base pairs which may well reduce the chances of frame shifting.

One cycle of the three-step elongation cycle has been completed, and the cycle is repeated until one of the three termination codons (stop codons) appears in the A-site.

Termination: There are no tRNA species that normally recognize stop codons. Instead, protein factors called release factors interact with these codons and cause release of the completed polypeptide chain. RFI recognizes the codons UAA and UAG, and RF2 recognizes UAA and UGA. RF3 helps either RFI or RF2 to carry out the reaction. The release factors make peptidyl transferase transfer the polypeptide to water rather than to the usual aminoacyl-tRNA, and thus the new protein is released. To remove the uncharged tRNA from the P-site and release the mRNA, EF-G together with ribosome release factor are needed for the complete dissociation of the subunits. IF_3 can now bind the small subunit to prevent inactive 70S ribosomes forming.

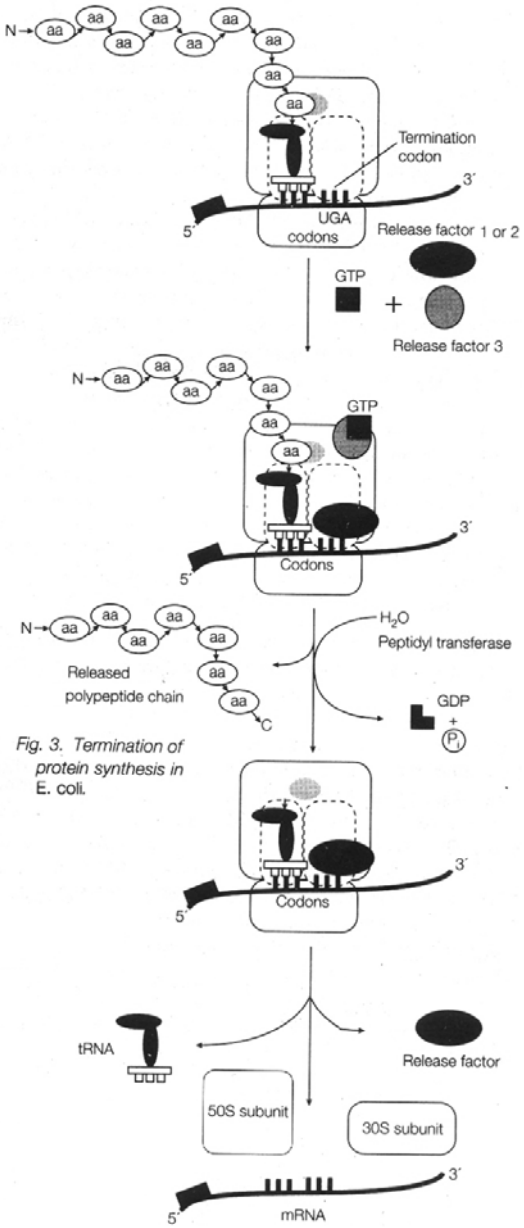


Fig. 3. Termination of protein synthesis in E. coli.

C5. MECHANISM OF PROTEIN SYNTHESIS IN EUKARYOTES

Apart from in the mitochondria and chloroplasts of eukaryotic cells (which are thought to originate from symbiotic prokaryotes; details of the mechanism of protein synthesis differ from that of prokaryotes. Most of these differences are in the initiation phase where a greater number of eIFs are involved. The method of finding the correct start codon involves a scanning process as there is no ribosome-binding sequence. Although there are two different tRNA species for methionine, one of which is the initiator tRNA, the attached methionine does not become converted to N-formylmethionine. A comparison of the factors involved in prokaryotes and eukaryotes is given in *Table.*

Prokaryotic	Eukaryotic	Function
Initiation factors		
IF1.IF3	e1F3, eIF4C, eIF6	Binding to ribosome subunits
	eIF4B, eIF4F	Binding to mRNA
IF2	e1F2. eIF2B	Initiator tRNA delivery
	eIF5	Displacement of other factors
Elongation factors		
EF-Tu	eEF1α	Aminoacyl tRNA delivery to ribosome
EF-Ts	eEF1$\beta\gamma$	Recycling of EF-Tu or eEF1a
EF-G	eEF2	Translocation
Termination factors		
RF1/ RF2/ RF3	eRF	Polypeptide chain release

Scanning:

Since there is no Shine-Dalgarno sequence in eukaryotic mRNA, the mechanism of selecting the start codon must be different. Kozak proposed a scanning hypothesis in which the 40S subunit, already containing the initiator tRNA, attaches to the 5' -end of the mRNA and scans along the mRNA until it finds an appropriate AUG. This is not always the first one as it must be in the correct sequence context (5' -CCRCCAUGG-3'), where R = purine.

Initiation:

Figure shows the steps and factors involved in the initiation stage of protein synthesis in eukaryotes. Although there are at least nine reasonably well-defined initiation factors involved in eukaryotic protein synthesis, some have analogous functions to the three prokaryotic IFs. They can be grouped according to their functions as follows:

- those binding to ribosomal subunits, such as e1F6, eIF3 and eIF4C;
- those binding to the mRNA to recognize the 5' -cap and to melt secondary structure, such as eIF4B and eIF4F, which is a complex of a cap-binding protein, eIF4A and eIF4E;
- those involved in initiator tRNA delivery, such as eIF2 and eIF2B;
- those that displace other factors, such as eIF5 which releases two other factors so the 60S subunit can bind.

The following events take place, starting with a free 40S subunit and a 5'capped mRNA molecule. eIF3 and 4C bind to the 405 subunit which allows it to bind a complex of three components (**ternary complex**) - the initiator tRNA, eIF2 and GTP.

Note this different order of assembly in eukaryotes where the initiator tRNA is bound to the small subunit before the mRNA binds. Before this large complex can bind to the mRNA, the latter must have interacted with eIF4B and 4F (which recognizes the 5' -cap) and, using energy from A TP, have been unwound to remove secondary structure. When the 40 S subunit complex has bound to the mRNA complex via the 5'cap, ATP is used as the mRNA is scanned to find the AUG start codon. This is usually the first one. To allow the 60S subunit to bind, eIF5 must displace eIF2 and e1F3, and GTP is hydrolyzed. eIF4C is released when it has assisted 605 subunit binding to form the complete 80S **initiation complex**.

The released eIF2.GDP complex is recycled by eIF2B and the rate of recycling (and hence the rate of initiation of protein synthesis) is regulated by phosphorylation of the a-subunit of eIF2.

Certain events, such as viral infection and the resultant production of interferon, cause an inhibition of protein synthesis by promoting **phosphorylation of eIF2**.

Elongation:

The protein synthesis elongation cycle in prokaryotes and eukaryotes is quite similar. Three factors are required with properties similar to their prokaryotic counterparts *(Table*). eEFlα, eEF1$\beta\gamma$ and eEF2 have the roles described for EFTu, EF-Ts and EF-G respectively.

In eukaryotes, a single release factor, eRF, recognizes all three stop codons and performs the roles carried out by RFI (or RF2) plus RF3 in prokaryotes. eRF requires GTP for activity, but it is not yet clear whether there is a eukaryotic equivalent of ribosome release factor required for dissociation of the subunits from the mRNA.

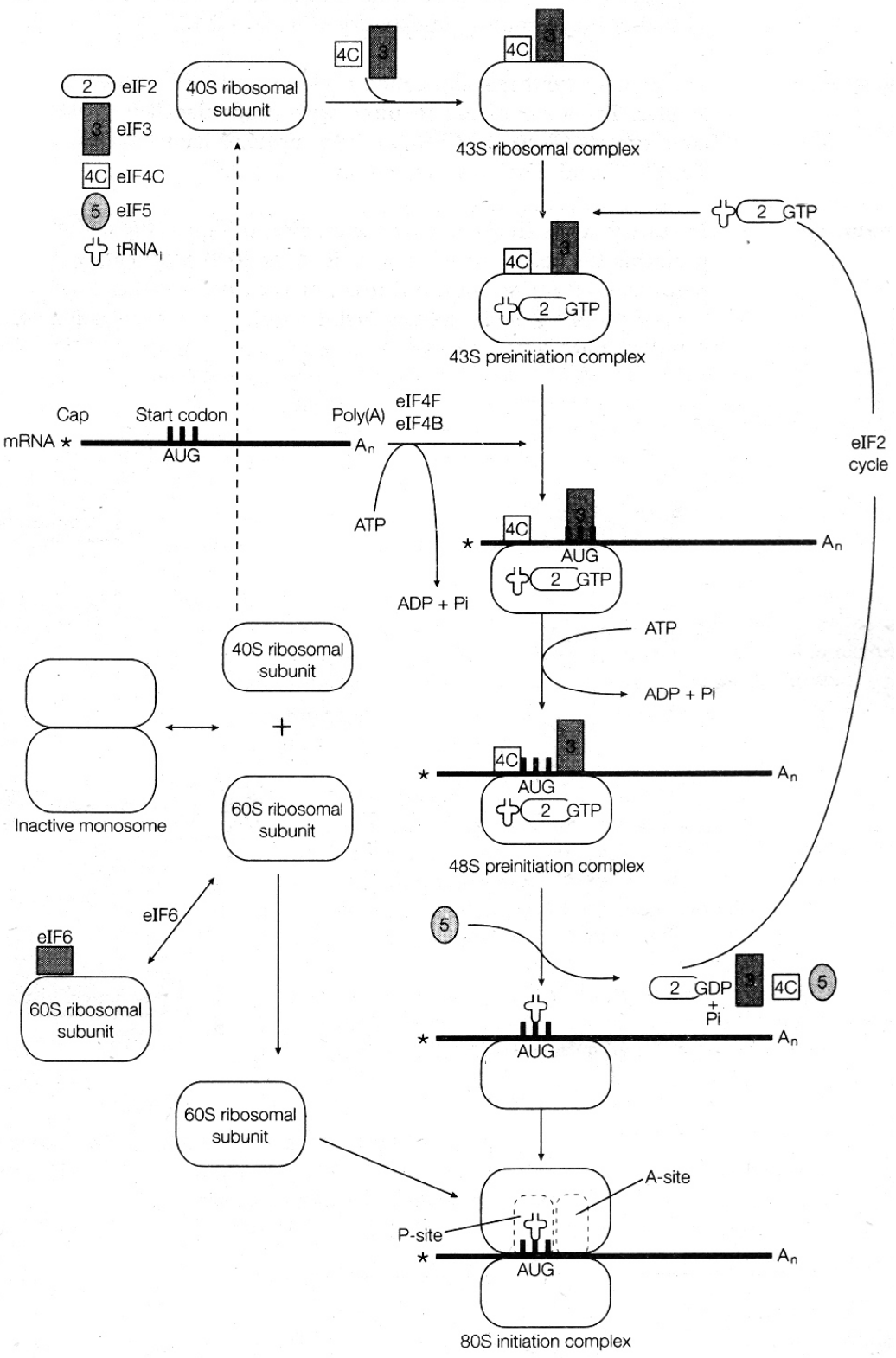


Fig. 1. Initiation of protein synthesis in eukaryotes.

C6. PROOF READING AND REGULATION OF PROTEIN SYNTHESIS

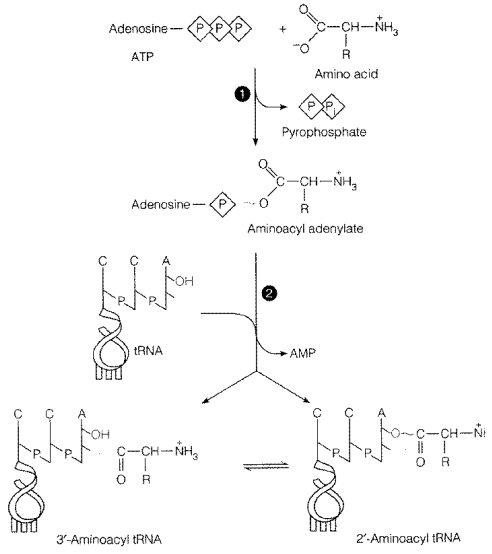
6.1 Proof reading of Translation:

Two sources of errors during translation:

- Attachment of an incorrect amino acid to a tRNA
- Mispairing of the tRNA anticodon with the mRNA codon

Two proofreading mechanisms exist to prevent these errors:

- Proofreading *before* aminoacyl adenylate intermediate is attached to tRNA.
- Kinetic proofreading *before* peptide bond formation: A delay is introduced between the binding of an aminoacyl-tRNA to the codon and the formation of the peptide bond to allow errors to be corrected:
  - EF-Tu-GTP binds an aminoacyl-tRNA and bring it into the A-site.
  - EF-Tu allows the anticodon to interact with the codon but prevents peptide bond formation.
  - An incorrect tRNA will bind weakly to the codon and will dissociate from the codon before an incorrect amino acid is incorporated into the polypeptide.
  - Correct codon-anticodon matching triggers hydrolysis of GTP by the EF-Tu, after which EF-Tu-GDP dissociates.
  - Peptide bond formation proceeds.



6.2 Constitutive control of protein synthesis.

Constitutive levels of protein synthesis are dependent upon mRNA structure and reflect such factors as mRNA stability, the sequence and context of the ribosome binding site, the presence of secondary structure, the choice of initiation codon, and codon bias throughout the open reading frame.

Secondary structure and the choice of termination codon also permit a group of regulatory phenomena collectively described as programmed misreading, where the normal interpretation of the sequence of codons is suppressed. Examples of programmed misreading include read through, selenocysteine insertion, and frameshifting. Read-through occurs at weak termination codons (i.e. termination codons where release factors are limiting), and there is competition between release factors and tRNAs with tolerable anticodon sequences. Selenocysteine insertion occurs at the termination codon UGA, and requires a secondary structure, the *selenocysteine insertion sequence*.

Frameshifting or recoding occurs when the ribosome pauses at a secondary structure or rare codon, shifts forwards or backwards by a single nucleotide, and continues translation in a different reading frame. This can be used to change the reading frame midway through translation (e.g. in the retroviral *pol* gene) to induce early truncation (e.g. the *MS2* lysis gene) or to prevent truncation and extend a gene product (e.g. in the *E. coli dnaX* gene).

In bacteria, the operon environment allows a novel form of translational regulation where the translation of downstream genes is dependent on the translation of upstream genes in a polycistronic message. The spacing of the individual open reading frames is important. Where the space between successive open reading frames is greater than approximately 30 nucleotides, translation begins with discrete initiation events and each locus is independent. If there is a shorter gap, the ribosome can reinitiate by bridging the open reading frames, i.e. without first dissociating from the template. The initiation signal is different from the normal SD-sequence, and spontaneous assembly of ribosome subunits in the typical manner occurs with low efficiency. A mutation which blocks or interrupts translation of the upstream gene therefore also effects translation of the downstream gene; this is termed a polar mutation (e.g. *lac operon).*

The translation of eukaryotic mRNA is usually cap-dependent and facilitated by scanning for the first initiator codon. The picornavirus family provides an exception, in that the genomes contain internal ribosome entry sites (IRES), i.e. motifs which form secondary structures allowing internal initiation of protein synthesis. Internal initiation is not dependent upon the cap-binding protein eIF- 4F, and occurs when this protein is inactive (the *picornaviridae* block host protein synthesis by inhibiting eIF-4F as part of their infection strategy). However, no viral proteins are required for IRES function, i.e. initiation is dependent upon other host initiation factors. Internal initiation may also occur for certain cellular transcripts, including *Drosophila Antennapedia* and

mouse *fgf2.* The abundance of IRES motifs and their significance in the control of gene expression is unknown; they have been defined in functional terms and appear to share no conserved structural features.

Global regulation of protein synthesis

The components of the 'basal apparatus' of protein synthesis (the initiation, elongation and termination factors, and the components of the ribosome) may be regulated and may exercise a global control over protein synthesis. For example, several eukaryotic viruses shut down host protein synthesis by phosphorylating the initiation factor eIF-2, and this strategy may also be used by the cell itself where global repression of protein synthesis is required (e.g. when cells are subjected to heat shock). Phosphorylation of eIF-2 in *S. cerevisiae* allows leaky scanning, where the ribosome can skip weak initiators and bind to strong ones. This lifts translational repression of genes such as *GCN4,* where there are several unproductive AUG codons between the cap and the definitive start codon on the mRNA. When eIF-2 is unphosphorylated, the ribosome attempts to initiate protein synthesis at the first AUG, which is followed by an in-frame termination codon.

Narrow domain regulation of protein synthesis

The translation of specific mRNAs can also be regulated individually. Well-characterized examples include the mammalian ferritin mRNA and ribosomal protein L5 mRNA in *E. coli.*

The efficiency of protein synthesis is under constitutive control, but may also be regulated either globally or at the level of individual transcripts. In the ferritin system, translation of ferritin mRNA is blocked when the concentration of intracellular iron is low. The inhibition depends upon iron-response elements *(IRES)* in the transcript, and is mediated by IRE-binding protein, which is inactivated in the presence of iron. The ferritin IRE is present in the 5' untranslated region of the transcript and IRE-BP binding inhibits protein synthesis by preventing ribosome scanning. In the ferritin mRNA, IREs are not associated with AU rich mRNA instability sites and do not affect mRNA turnover, as is the case for transferrin receptor mRNA.

The L5 system represents a feedback control for the synthesis of ribosomal components. Ribosomal protein S8 binds to a stem loop structure formed by residues 588-651 of the 16s rRNA. A similar structure is formed by the 5' untranslated region and the first 30 bases of the coding region of the W ribosomal protein mRNA. Protein S8 binds to L5 mRNA (albeit with lower efficiency than does 165 rRNA) and inhibits protein synthesis. Excess 16s rRNA sequesters all the S8 protein and thus derepresses L5 translation, whereas if 16s rRNA is limiting (and there is therefore no need for L5 translation), the excess S8 protein prevents L5 translation.

A final example of specific translational regulation is provided by antisense *RNA*. This is found mainly in prokaryote systems (e.g. in the control of *Q* protein synthesis in bacteriophage λ, in the control of plasmid replication genes, and in the control of transposase synthesis in various transposons, but also in eukaryotes (e.g. in the control of *Fgf*-2 synthesis in the chick limb bud). More recently, antisense suppression of translation has been used as a strategy to block gene function without *in vitro* or targeted mutagenesis allowing the effects of null phenotypes to be determined without gene disruption.

Table: Antibiotics as an inhibitor of Protein Synthesis:

Inhibitor	Subunit or Factor Affected	Step(s) Blocked	Reaction Affected	Cell Type Affected
Aurintricarboxylic acid	30S/40S	Initiation	Binding of m-RNA	Prokaryotes/Eukaryotes
Chloramphenicol	50 S	Elongation	Peptide bond Formation	Prokaryotes
Colcicin E3	30S	Initiation	Binding of m-RNA	
		Elongation	Binding of aminoacyl t-RNA	Prokaryotes
Cyclohexamide	60S	Initiation	Binding of Initiator t-RNA	
		Elongation	Translocation (t-RNA release from P Site)	Eukaryotes
Diptheria Toxin	eEF-2	Elongation	Translation	
Erthromycin	50 S	Initiation	Formation of Initiation complex	
Fusidic Acid	EF-G/eEF-2	Elongation	Translocation	Eukaryotes
Kasugamycin	30 S	Initiation	Binding of initiator t-RNA	Prokaryotes
Puromycin	50 S / 60S	Elongation	Peptide bond formation (Trigger immature chain release)	Prokaryotes/ Eukaryotes Prokaryotes
Spectinomycin	30 S	Elongation	Translocation	Prokaryotes/ Eukaryotes
Streptomycin	30 S	Initiation	Binding if Initiator t-RNA	Prokaryotes
		Elongation	Binding of aminoacyl t-RNA (Induced Misreading)	
Tetracycline	30 S	Elongation	Binding of aminoacyl t-RNA	Prokaryotes
		Termination	Binding of RF-1 And RF-2	Prokaryotes

C7. POST TRANSLATIONAL EVENTS:

A polypeptide chain is synthesized by a complex process called **translation** in which the assembly of amino acids in a particular sequence is dictated by **messenger RNA (mRNA)**.

We describe how the cell promotes the proper folding of a nascent polypeptide chain and, in many cases, modifies residues or cleaves the polypeptide backbone to generate the final protein. In addition, the cell has error-checking processes that eliminate incorrectly synthesized or folded proteins. Incorrectly folded proteins usually lack biological activity and, in some cases, may actually be associated with disease. Protein misfolding is suppressed by two distinct mechanisms. First, cells have systems that reduce the chances for misfolded proteins to form. Second, any misfolded proteins that do form, as well as cytosolic proteins no longer needed by a cell, are degraded by a specialized cellular garbage-disposal system.

The Information for Protein Folding Is Encoded in the Sequence

Any polypeptide chain containing *n* residues could, in principle, fold into 8^n conformations. This value is based on the fact that only eight bond angles are stereochemically allowed in the polypeptide backbone. In general, however, all molecules of any protein species adopt a single conformation, called the *native state;* for the vast majority of proteins, the native state is the most stably folded form of the molecule.

What guides proteins to their native folded state? The answer to this question initially came from in vitro studies on protein refolding. Thermal energy from heat, extremes of pH that alter the charges on amino acid side chains, and chemicals such as urea or guanidine hydrochloride at concentrations of 6–8 M can disrupt the weak noncovalent interactions that stabilize the native conformation of a protein. The **denaturation** resulting from such treatment causes a protein to lose both its native conformation and its biological activity.

Many proteins that are completely unfolded in 8 M urea and β-mercaptoethanol (which reduces disulfide bonds) spontaneously *renature* (refold) into their native states when the denaturing reagents are removed by dialysis. Because no cofactors or other proteins are required, in vitro protein folding is a self directed process.

In other words, sufficient information must be contained in the protein's primary sequence to direct correct refolding. The observed similarity in the folded, three dimensional structures of proteins with similar amino acid sequences, provided other evidence that the primary sequence also determines protein folding in vivo.

Folding of Proteins in Vivo Is Promoted by Chaperones

Although protein folding occurs in vitro, only a minority of unfolded molecules undergo complete folding into the native conformation within a few minutes. Clearly, cells require a faster, more efficient mechanism for folding proteins into their correct shapes; otherwise, cells would waste much energy in the synthesis of nonfunctional proteins and in the degradation of misfolded or unfolded proteins. Indeed, more than 95 percent of the proteins present within cells have been shown to be in their native conformation, despite high protein concentrations (200–300 mg/ml), which favor the precipitation of proteins in vitro.

The explanation for the cell's remarkable efficiency in promoting protein folding probably lies in **chaperones,** a class of proteins found in all organisms from bacteria to humans. Chaperones are located in every cellular compartment, bind a wide range of proteins, and function in the general protein-folding mechanism of cells.

Two general families of chaperones are reconized:

- *Molecular chaperones,* which bind and stabilize unfolded or partly folded proteins, thereby preventing these proteins from aggregating and being degraded
- *Chaperonins,* which directly facilitate the folding of proteins

Molecular chaperones consist of Hsp70 and its homologs: Hsp70 in the cytosol and mitochondrial matrix, BiP in the endoplasmic reticulum, and DnaK in bacteria. First identified by their rapid appearance after a cell has been stressed by heat shock, Hsp70 and its homologs are the major chaperones in all organisms. (Hsc70 is a constitutively expressed homolog of Hsp70.) When bound to ATP, Hsp70-like proteins assume an open form in which an exposed hydrophobic pocket transiently binds to exposed hydrophobic regions of the unfolded target protein. Hydrolysis of the bound ATP causes molecular chaperones to assume a closed form in which a target protein can undergo folding. The exchange of ATP for ADP releases the target protein (Figure a, *top*). This cycle is speeded by the co-chaperone Hsp40 in eukaryotes.

In bacteria, an additional protein called GrpE also interacts with DnaK, promoting the exchange of ATP for the bacterial co-chaperone DnaJ and possibly its dissociation. Molecular chaperones are thought to bind all nascent polypeptide chains as they are being

synthesized on ribosomes. In bacteria, 85 percent of the proteins are released from their chaperones and proceed to fold normally; an even higher percentage of proteins in eukaryotes follow this pathway.

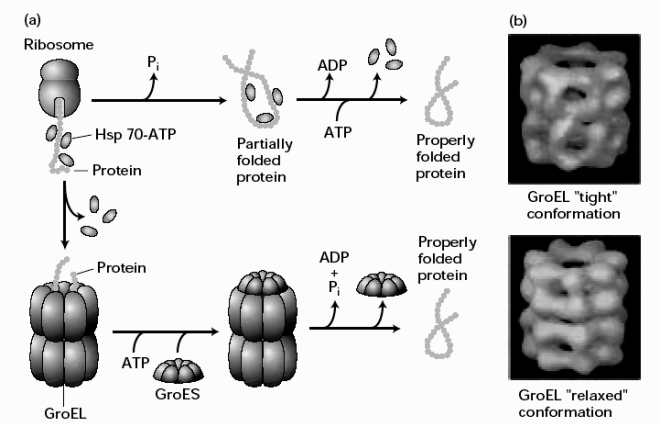


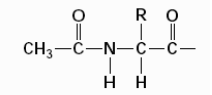
FIGURE Chaperone- and chaperonin-mediated protein folding.
(a) Many proteins fold into their proper three- dimensional structures with the assistance of Hsp70-like proteins. These molecular chaperones transiently bind to a nascent polypeptide as it emerges from a ribosome. Proper folding of other proteins depends on chaperonins such as the prokaryotic GroEL, a hollow, barrel-shaped complex of 14 identical 60,000-MW subunits arranged in two stacked rings. One end of GroEL is transiently blocked by the co- chaperonin GroES, an assembly of 10,000-MW subunits. (b) In the absence of ATP or presence of ADP, GroEL exists in a "tight" conformational state that binds partly folded or misfolded proteins. Binding of ATP shifts GroEL to a more open, "relaxed" state, which releases the folded protein.

The proper folding of a large variety of newly synthesized or translocated proteins also requires the assistance of chaperonins. These huge cylindrical macromolecular assemblies are formed from two rings of oligomers. The eukaryotic chaperonin *TriC* consists of eight subunits per ring. In the bacterial, mitochondrial, and chloroplast chaperonin, known as *GroEL,* each ring contains seven identical subunits (Figure b). The GroEL folding mechanism, which is better understood than TriC-mediated folding, serves as a general model (Figure a, *bottom*). In bacteria, a partly folded or misfolded polypeptide is inserted into the cavity of GroEL, where it binds to the inner wall and folds into its native conformation. In an ATP-dependent step, GroEL undergoes a conformational change and releases the folded protein, a process assisted by a co-chaperonin, GroES, which caps the ends of GroEL.

Many Proteins Undergo Chemical Modification of Amino Acid Residues

Nearly every protein in a cell is chemically modified after its synthesis on a ribosome. Such modifications, which may
alter the activity, life span, or cellular location of proteins, entail the linkage of a chemical group to the free $-NH_2$ or –COOH group at either end of a protein or to a reactive side chain group in an internal residue. Although cells use the 20 amino acids shown in Figure 2 to synthesize proteins, analysis of cellular proteins reveals that they contain upward of 100 different amino acids. Chemical modifications after synthesis account for this difference.

Acetylation, the addition of an acetyl group (CH_3CO) to the amino group of the N-terminal residue, is the most common form of chemical modification, affecting an estimated 80 percent of all proteins:



This modification may play an important role in controlling the life span of proteins within cells because non acetylated proteins are rapidly degraded by intracellular proteases. Residues at or near the termini of some membrane proteins are chemically modified by the addition of long lipid like groups. The attachment of these hydrophobic "tails," which function to anchor proteins to the lipid bilayer, constitutes one way that cells localize certain proteins to membranes.

Acetyl groups and a variety of other chemical groups can also be added to specific internal residues in proteins (Figure 2). An important modification is the *phosphorylation* of serine, threonine, tyrosine, and histidine residues. We will encounter numerous examples of proteins whose activity is regulated by reversible phosphorylation and dephosphorylation. The side chains of asparagine, serine, and threonine are sites for *glycosylation,* the attachment of linear and branched carbohydrate chains. Many secreted proteins and membrane proteins contain glycosylated residues. Other post-translational modifications found in selected proteins include the *hydroxylation* of proline and lysine residues in collagen, the *methylation* of histidine residues in membrane receptors, and the *γ-carboxylation* of glutamate in prothrombin, an essential blood-clotting factor. A special modification, discussed shortly, marks cytosolic proteins for degradation.

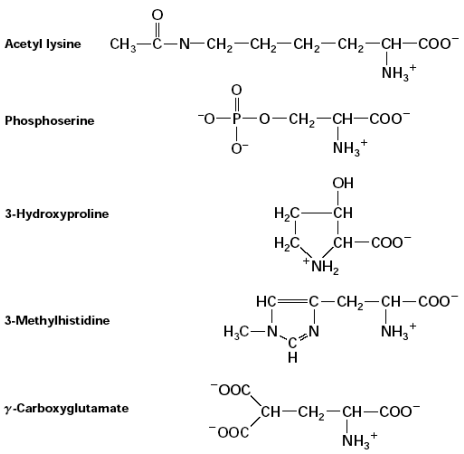


FIGURE Common modifications of internal amino acid residues found in proteins.

Peptide Segments of Some Proteins Are Removed After Synthesis

After their synthesis, some proteins undergo irreversible changes that do not entail changes in individual amino acid residues. This type of post-translational alteration is sometimes called *processing.* The most common form is enzymatic cleavage of a backbone peptide bond by proteases, resulting in the removal of residues from the C- or N-terminus of a polypeptide chain. Proteolytic cleavage is a common mechanism for activating enzymes that function in blood coagulation, digestion, and programmed cell death.

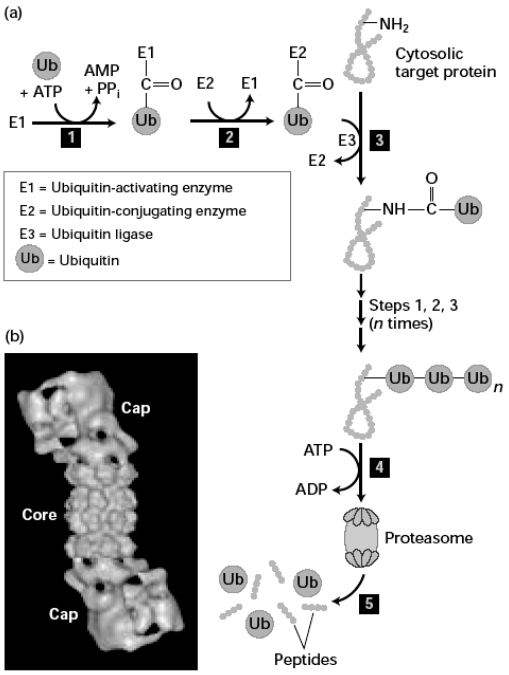
Proteolysis also generates active peptide hormones, such as EGF and insulin, from larger precursor polypeptides. An unusual and rare type of processing, termed *protein self-splicing,* takes place in bacteria and some eukaryotes. This process is analogous to editing film: an internal segment of a polypeptide is removed and the ends of the polypeptide are rejoined. Unlike proteolytic processing, protein self-splicing is an autocatalytic process, which proceeds by itself without the participation of enzymes. The excised peptide appears to eliminate itself from the protein by a mechanism similar to that used in the processing of some RNA molecules. In vertebrate cells, the processing of some proteins includes self-cleavage, but the subsequent ligation step is absent. One such protein is Hedgehog, a membrane bound signaling molecule that is critical to a number of developmental processes.

Ubiquitin Marks Cytosolic Proteins for Degradation in Proteasomes

In addition to chemical modifications and processing, the activity of a cellular protein depends on the amount present, which reflects the balance between its rate of synthesis and rate of degradation in the cell. Here we examine protein degradation, focusing on the major pathways for degrading cytosolic proteins. The life span of intracellular proteins varies from as short as a few minutes for mitotic cyclins, which help regulate passage through mitosis, to as long as the age of an organism for proteins in the lens of the eye. Eukaryotic cells have several intracellular proteolytic pathways for degrading misfolded or denatured proteins, normal proteins whose concentration must be decreased, and extracellular proteins taken up by the cell. One major intracellular pathway is degradation by enzymes within **lysosomes,** membrane-limited organelles whose acidic interior is filled with hydrolytic enzymes. Lysosomal degradation is directed primarily toward extracellular proteins taken up by the cell and aged or defective organelles of the cell.

Distinct from the lysosomal pathway are cytosolic mechanisms for degrading proteins. Chief among these mechanisms is a pathway that includes the chemical modification of a lysine side chain by the addition of **ubiquitin,** a 76-residue polypeptide, followed by degradation of the ubiquitin-tagged protein by a specialized proteolytic machine. Ubiquitination is a three-step process (Figure 3a):

- Activation of *ubiquitin-activating enzyme* (E1) by the addition of a ubitiquin molecule, a reaction that requires ATP
- Transfer of this ubiquitin molecule to a cysteine residue in *ubiquitin-conjugating enzyme* (E2)
- Formation of a peptide bond between the ubiquitin molecule bound to E2 and a lysine residue in the target protein, a reaction catalyzed by *ubiquitin ligase* (E3)



▲ **FIGURE** Ubiquitin-mediated proteolytic pathway.

This process is repeated many times, with each subsequent ubiquitin molecule being added to the preceding one. The resulting polyubiquitin chain is recognized by a **proteasome,** another of the cell's molecular machines (Figure 3b). The numerous proteasomes dispersed throughout the cell cytosol proteolytically cleave ubiquitin-tagged proteins in an ATP dependent process that yields short (7- to 8-residue) peptides and intact ubiquitin molecules.

Cellular proteins degraded by the ubiquitin-mediated pathway fall into one of two general categories: (1) native cytosolic proteins whose life spans are tightly controlled and (2) proteins that become misfolded in the course of their synthesis in the endoplasmic reticulum (ER). Both contain sequences recognized by the ubiquitinating enzyme complex. The cyclins, for example, are cytosolic proteins whose amounts are tightly controlled throughout the cell cycle. These proteins contain the internal sequence Arg-X-X-Leu-Gly-X-Ile-Gly-Asp/Asn (X can be any amino acid), which is recognized by specific ubiquitinating enzyme complexes. At a specific time in the cell cycle, each

cyclin is phosphorylated by a cyclin kinase. This phosphorylation is thought to cause a conformational change that exposes the recognition sequence to the ubiquitinating enzymes, leading to degradation of the tagged cyclin. Similarly, the misfolding of proteins in the endoplasmic reticulum exposes hydrophobic sequences normally buried within the folded protein. Such proteins are transported to the cytosol, where ubiquitinating enzymes recognize the exposed hydrophobic sequences.

The immune system also makes use of the ubiquitin mediated pathway in the response to altered self-cells, particularlyvirus-infected cells. Viral proteins within the cytosol of infected cells are ubiquitinated and then degraded in proteasomes specially designed for this role. The resulting antigenic peptides are transported to the endoplasmic reticulum, where they bind to class I major histocompatibility complex (MHC) molecules within the ER membrane. Subsequently, the peptide-MHC complexes move to the cell membrane where the antigenic peptides can be recognized by cytotoxic T lymphocytes, which mediate the destruction of the infected cells.

Digestive Proteases Degrade Dietary Proteins

The major extracellular pathway for protein degradation is the system of digestive proteases that breaks down ingested proteins into peptides and amino acids in the intestinal tract. Three classes of proteases function in digestion. *Endoproteases* attack selected peptide bonds within a polypeptide chain. The principal endoproteases are pepsin, which preferentially cleaves the backbone adjacent to phenylalanine and leucine residues, and trypsin and chymotrypsin, which cleave the backbone adjacent to basic and aromatic residues. *Exopeptidases* sequentially remove residues from the N-terminus (aminopeptidases) or C-terminus (carboxypeptidases) of a protein. *Peptidases* split oligopeptides containing as many as about 20 amino acids into di- and tripeptides and individual amino acids. These small molecules are then transported across the intestinal lining into the bloodstream. To protect a cell from degrading itself, endoproteases and carboxypeptidases are synthesized and secreted as inactive forms (zymogens): pepsin by chief cells in the lining of the stomach; the others by pancreatic cells. Proteolytic cleavage of the zymogens within the gastic or intestinal lumen yields the active enzymes. Intestinal epithelial cells produce aminopeptidases and the di- and tripeptidases.

D1. REGULATION OF GENE EXPRESSION IN PROKARYOTES (OPERON CONCEPT)

Jacob and Monod proposed the operon model in 1961 for the co-ordinate regulation of transcription of genes involved in specific metabolic pathways. The operon is a unit of gene expression and regulation which typically includes:

- The **structural genes** (any gene other than a regulator) for enzymes involved in a specific biosynthetic pathway whose expression is coordinately controlled.
- **Control elements** such as an operator sequence, which is a DNA sequence that regulates transcription of the structural genes.
- **Regulator gene(s)** whose products recognize the control elements, for example a repressor which binds to and regulates an operator sequence.

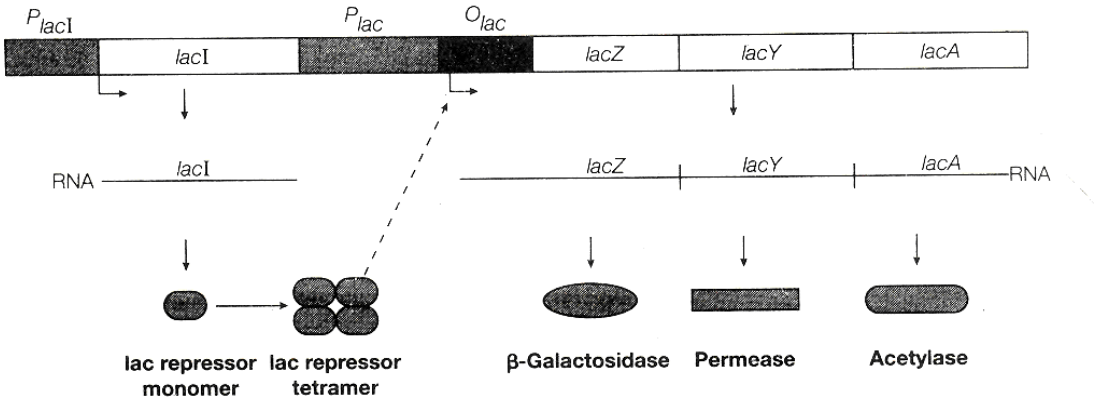


Fig. **1** *Structure of the lactose operon.*

1.1 The Lactose Operon: *Escherichia coli* can use lactose as a source of carbon. The enzymes required for the use of lactose as a carbon source are only synthesized when lactose is available as the sole carbon source. The lactose operon (or *lac* operon) consists of three structural genes: *lacZ,* which codes *for* β-galactosidase, an enzyme responsible for hydrolysis of lactose to galactose and glucose; *lacY* which encodes a galactoside permease which is responsible for lactose transport across the bacterial

cell wall; and *lacA,* which encodes a thiogalactoside transacetylase. The three structural genes are encoded in a single transcription unit *lacZYA*, which has a single promoter, *P lac.*

This organization means that the three lactose operon structural proteins are expressed together as a polycistronic mRNA containing more than one coding region under the same regulatory control. The *lacZYA* transcription unit contains an operator site *O lac* which is positioned between bases -5 and +21 at the 5' -end of the *P lac* promoter region. This site binds a protein called the lac repressor which is a potent inhibitor of transcription when it is bound to the operator. The lac repressor is encoded by a separate regulatory gene *lac I* which is also a part of the lactose operon; *lac I* is situated just upstream from *P lac.*

The Lac Repressor: The *lacI* gene encodes the lac repressor, which is active as a tetramer of identical subunits. It has a very strong affinity *for* the *lac* operator-binding site, *olac* and also has a generally high affinity for DNA. The *lac* operator site consists of 28 bp which is palindromic. (A palindrome has the same DNA sequence when one strand is read left to right in a 5' to 3' direction and the complementary strand is read right to left in a 5' to 3' direction. This inverted repeat symmetry of the operator matches the inherent symmetry of the lac repressor which is made up of four identical subunits.

In the absence of lactose, the repressor occupies the operator-binding site. It seems that both the lac repressor and the RNA polymerase can bind simultaneously to the *lac* promoter and operator sites. The lac repressor actually increases the binding of the polymerase to the *lac* promoter by two orders of magnitude

Induction: It means that when glucose is absent and lactose is present lac repressor is cannot bind to the *Olac* operator DNA sequence. RNA polymerase is also likely to be bound to the *Plac* promoter sequence. In the presence of lactose permease allows its uptake, and **β**-galactosidase catalyzes the conversion of some lactose to allolactose. Allolactose acts as an inducer and binds to the lac repressor. This causes a change in the conformation of the repressor tetramer, reducing its affinity for the *lac* operator. The removal of the lac repressor from the operator site allows the polymerase (which is already sited at the adjacent promoter) to rapidly begin transcription of the *lacZYA* genes. Thus, the addition of lactose, or a synthetic inducer such as isopropyl β-o-thiogalactopyranoside (IPTG), very rapidly stimulates transcription of the lactose operon structural genes. The subsequent removal of the inducer leads to an almost immediate inhibition of this induced transcription, since the free lac repressor rapidly re occupies the operator site and the *lacZYA* RNA transcript is extremely unstable.

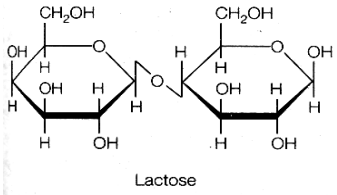
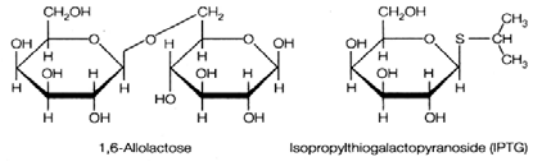


Fig. 2. Structures of lactose, allolactose



and IPTG.

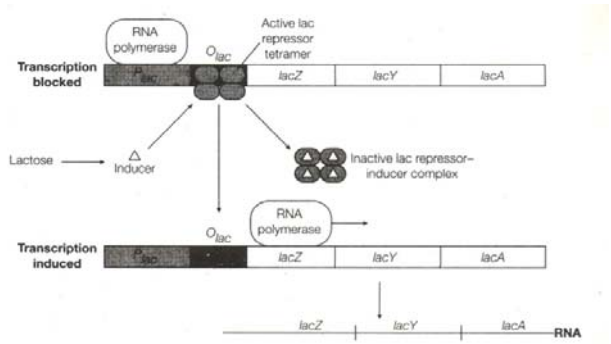


Fig. 3. Binding of inducer inactivates the lac repressor.

cAMP Receptor Protein: The *Plac* promoter is not a strong promoter. *Plac* and related promoters do not have strong -35 sequences and some even have weak -10 consensus sequences. For high level transcription, they require the activity of a specific activator protein called cAMP receptor protein (CRP). CRP may also be called catabolite activator protein or CAP. When glucose is present, E. *coli* does not require alternative carbon Sources such as lactose.

CRP which exists as a dimer which cannot bind to DNA on its own, nor regulate late transcription. Glucose reduces the level of cAMP in the cell. When glucose is absent, the levels of cAMP in E. *coli* increase and CRP binds to cAMP forming CRP-AMP complex which binds to the lactose operon promoter *Plac* just upstream from the site for RNA polymerase. CRP binding induces a 90° bend in DNA and this is believed to enhance RNA polymerase binding to the promoter enhancing transcription by 50-fold. The CRP-binding site is an inverted repeat and may be adjacent to the promoter (as in the lactose operon), may lie within the promoter itself, or may be much further upstream from the promoter. Differences in the CRP-binding sites of the promoters of different catabolic operons may mediate different levels of response of these operons to cAMP *in vivo.*

1.2 Trp-Operon

The *trp* operon encodes five structural genes whose activity is required far tryptophan synthesis. The operon encodes a single transcription Unit which produces a 7 kb transcript which is synthesized downstream from the *trp* promoter and *trp* operator sites *Ptrp* and *Otrp.* Like many of the operons involved in amino acid biosynthesis, the *trp* operon has evolved systems for coordinated expression of these genes when the product of the biosynthetic pathway tryptophan is in short supply in the cell. As with the *lac* operon, the RNA product of this transcription unit is very unstable, enabling bacteria to respond rapidly to changing needs for tryptophan.

The Trp-Repressor: A gene product of the separate *trpR* operon, the trp repressor, specifically interacts with the operator site of the *trp* operon. The symmetrical operator sequence, which forms the trp repressor-binding site, overlaps with the *trp* promoter sequence between bases -21 and +3. The core binding site is a palindrome of 18 bp. The trp repressor binds tryptophan and can only bind to the operator when it is complexed with tryptophan. The repressor is a dimer of two subunits which have structural similarity to the CRP protein and lac repressor. The repressor dimer has a structure with a central core and two flexible DNA-reading heads each formed from the carboxyl-terminal half of one subunit. Only when tryptophan is bound to the repressor are the reading heads the correct distance apart, and the side chains in the correct conformation, to interact with successive major grooves of the DNA at the *trp* operator sequence. Tryptophan, the end-product of the enzymes encoded by the *trp* operon, therefore acts as a co-repressor and inhibits its own synthesis through end-product inhibition. The repressor reduces transcription initiation by around 70-fold. This is a much smaller transcriptional effect than that mediated by the binding of the lac repressor.

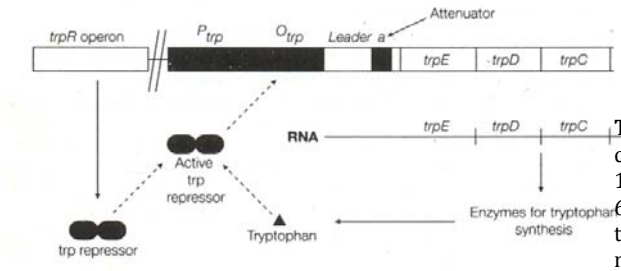


Fig. 1. Structure of the trp operon and function of the trp repressor.

Attenuator: At first, it was thought that the repressor was responsible for all of the transcriptional regulation of the *trp* operon. However, it was observed that the deletion of a sequence between the operator and the *trpE* gene coding region resulted in an increase in both the basal and the activated (derepressed) levels of transcription. This site is termed the attenuator and it lies towards the end of the transcribed leader sequence of 162 nt that precedes the *trpE* initiator codon. The attenuator is a rho-independent terminator site which has a short GC-rich palindrome followed by eight successive U residues. If this sequence is able to form a hairpin structure in the RNA transcript, then it acts as a highly strong terminator and only a 140 bp transcript is synthesized.

The Leader RNA structure: The leader sequence of the *trp* operon RNA contains four regions of complementary sequence which can form different base-paired RNA structures. These are termed sequences 1, 2, 3 and 4. The attenuator hairpin is the product of the base pairing of sequences 3 and 4 (3:4 structure). Sequences and 2 are also complementary and can form a second 1:2 hairpin. However sequence 2 is also complementary to sequence 3. If sequences 2 and 3 form 2:3 hairpin structure (**antitermination**), the 3:4 attenuator hairpin cannot be formed and transcription termination will not occur. Under normal conditions, the formation of the 1:2 and 3:4 hairpins is energetically favorable.

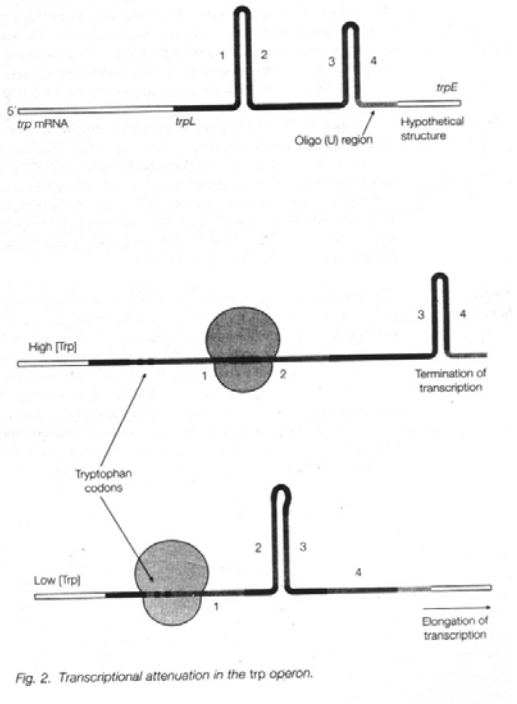


Fig. 2. Transcriptional attenuation in the trp operon.

The Leader Peptide: The leader RNA sequence contains an efficient ribosome-binding site and form a 14-amino-acid leader peptide encoded by bases 27-68 of the leader RNA, The 10th and 11 th codons of this leader peptide encode successive tryptophan residues, the end-product of the synthetic enzymes of the *trp* operon. This leader peptide has no obvious function as a polypeptide, and tryptophan is a rare amino acid therefore, the chances of two tryptophan codons in succession is low and under conditions of low tryptophan availability, the ribosome would be expected to pause at this site, The function of this leader peptide is to determine tryptophan availability and to regulate transcription termination.

Attenuation: Attenuation depends on the fact that transcription and translation are tight and are coupled in E. *coli;* translation can occur as an mRNA is being transcribed. 3'-end of the *trp* leader peptide

coding sequence overlaps complementary sequence 1; the two trp codons are within sequence 1 and the stop codon is between sequences 1 and 2. The availability of tryptophan (the ultimate product of the enzymes synthesized by the *trp* operon) is sensed through its being required in translation, and determines whether or not the terminator (3:4) hairpin forms in the mRNA. As transcription of the *trp* operon proceeds, the RNA polymerase pauses the end of sequence 2 until a ribosome begins to translate the leader peptide. Under conditions of high tryptophan availability, the ribosome rapidly incorporates tryptophan at the two trp codons and thus translates to the end of the leader message. The ribosome is then occluding sequence 2 and, as the polymerase reaches the terminator sequence, the 3:4 hairpin can form, and transcription may be terminated. This is the process of attenuation. Alternatively, if tryptophan is in scarce supply, it will not be available as aminoacyl tRNA for translation, and the ribosome will then pause at the two trp codons, occluding sequence 1. This leaves sequence 2 free to form a hairpin with sequence 3, known as the **anti-terminator terminator** (3:4) hairpin cannot form, and transcription continues in direction of *trp genes* and beyond. Thus the level of the end product, tryptophan, determines the probability that transcription will terminate early (attenuation), rather the proceeding through the whole operon.

The presence of tryptophan gives rise to a l0-fold repression of *trp* operon transcription through the process of attenuation alone. Combined with contribution by the *trp* repressor (70~fold), this means that tryptophan levels exert a 700-fold regulatory effect on expression from the *trp* operon. Attenuation occurs in at least six operons that encode enzymes concerned with amino acid biosynthesis. For example, the *His* operon has a leader which encodes a peptide with seven successive histidine codons. Not all of these other operons have the same combination of regulatory controls that are found in the *trp* operon. The *His* operon has no repressor-operator regulation, and attenuation forms the only mechanism of feedback control.

1.3 Transcriptional Regulation by Alternative σ Factors:

The $\alpha_2\beta\beta'w$ core enzyme of RNA polymerase is unable to start transcription at promoter sites. In order to specifically recognize the consensus -35 and -10 elements of general promoters, it requires the **σ** factor subunit. This subunit is only required for transcription initiation, being released from the core enzyme after initiation and before RNA elongation takes place. Thus, **σ** factors appear to be bifunctional proteins that simultaneously can bind to core RNA polymerase and recognize specific promoter sequences in DNA. Many bacteria, including *E. coli,* produce a set of **σ** factors that recognize different sets of promoters. Transcription initiation from single promoters or small groups of promoters is regulated commonly by single transcriptional repressors (such as the lac repressor) or transcriptional activators (such as the cAMP receptor protein, CRP). However, some environmental conditions require a massive change in the overall pattern of gene expression in the cell. Under such circumstances, bacteria may use a different set of **σ** factors to direct RNA polymerase binding to different promoter sequences. This process allows the diversion of the cell's basic transcription machinery to the specific transcription of different classes of genes.

Promoter recognition: The binding of an alternative **σ** factor to RNA polymerase can confer a new promoter specificity on the enzyme responsible for the general RNA synthesis of the cell. Comparisons of promoters activated by polymerase complexed to specific **σ** factors show that each **σ** factor recognizes a different combination of sequences centered approximately around the -35 and -to sites. It seems likely that **σ** factors themselves contact both of these regions, with the -to region J being most important. The σ^{70} subunit is the most common **σ** factor in E. *coli* which is responsible for recognition of general promoters which have consensus -35 and -10 elements.

Heat Shock: The response to heat shock is one example in E. *coli* where gene expression is altered significantly by the use of different **σ** factors. When E. *coli* is subjected to an increase in temperature, the synthesis of a set of around 17 proteins, called **heat-shock proteins**, is induced. If E. *coli* is transferred from 37 to 42°C, this burst of heat-shock protein synthesis is transient. However if the increase in temperature is more extreme, such as to 50°C, where growth of E. *coli* is not possible, then the heat-shock proteins are the only proteins synthesized. The promoters for E. *coli* heat-shock protein-encoding genes are recognized by a unique form of RNA polymerase holoenzyme containing a variant **σ** factor, σ^{32}, which is encoded by the *rpoH* gene. σ^{32} is a minor protein which is much less abundant than σ^{70}. Holoenzyme containing σ^{32} acts exclusively on promoters of heat-shock genes and does not recognize the general consensus promoters of most of the other genes *(Fig.*). Heat-shock promoters accordingly have different sequences to other general promoters which bind to σ^{70}.

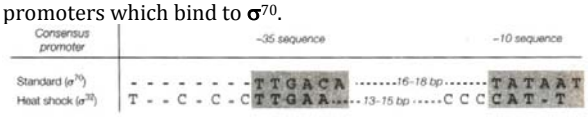


Fig. 1. Comparison of the heat-shock (α^{32}) and general (α^{70}) responsive promoters.

Sporulation in *Bacillus subtillis:* Vegetatively growing B. *subtilis* cells form bacterial spores in response to a sub-optimal environment. The formation of a spore (or sporulation) requires drastic changes in gene expression, including the cessation of the synthesis of almost all of the proteins required for vegetative existence as well as the production of proteins which are necessary for the resumption of' protein synthesis when the spore germinates under more optimal conditions. The process of spore formation involves the asymmetrical division of the bacterial cell into two compartments, the forespore, which forms the spore, and the mother cell, which is

eventually discarded. This system is considered one of the most fundamental examples of cell differentiation. The RNA polymerase in *B. subtilis* is functionally identical to that in E. *coli.* The vegetatively growing *B. subtilis* contains a diverse set of σ factors. Sporulation is regulated by a further set of σ factors in addition to those of the vegetative cell. Different σ factors are specifically active before cell partition occurs, in the forespore and in the mother cell. Cross-regulation of this compartmentalization permits the forespore and mother cell to tightly co-ordinate the differentiation process.

Bacteriophage σ factors: Some bacteriophages provide new σ subunits to endow the host RNA polymerase with a different promoter specificity and hence to selectively express their own phage genes (e.g. phage T4 in *E. coli* and SP01 in *B. subtilis).* This strategy is an effective alternative to the need for the phage to encode its own complete polymerase (e.g. bacteriophage T_7). The *B. subtilis* bacteriophage SP01 expresses a 'cascade' of σ factors in sequence to allow its own genes to be transcribed at specific stages during virus infection. Initially, **early genes** are expressed by the normal bacterial holoenzyme. Among these **early genes** is the gene encoding σ^{28}, which then displaces the bacterial σ factor from the RNA polymerase. The σ^{28}-containing holoenzyme is then responsible for expression of the **middle genes**. The phage middle genes include genes 33 and 34 which specify a further σ factor that is responsible for the specific transcription of **late genes**. In this way, the bacteriophage uses the host's RNA polymerase machinery and expresses its genes in a defined sequential order.

D2. REGULATION OF GENE EXPRESSION IN EUKARYOTES

2. 1. Chromatin Modifications (Hetero-chromatization) and Genome Expression

Chromatin is the complex of genomic DNA and chromosomal proteins present in the eukaryotic nucleus. Chromatin structure is hierarchic, ranging from the two lowest levels of DNA packaging —the nucleosome and the 30 nm chromatin fiber to the metaphase chromosomes, which represent the most compact form of chromatin in eukaryotes and occur only during nuclear division. After division, the chromosomes become less compact and cannot be distinguished as individual structures. When non-dividing nuclei are examined by light microscopy all that can be seen is a mixture of lightly and darkly staining areas within the nucleus. The dark areas, which tend to be concentrated around the periphery of the nucleus, are called heterochromatin and contain DNA that is still in a relatively compact organization, although still less compact than in the metaphase structure. Two types of heterochromatin are recognized:

1. *Constitutive heterochromatin* is a permanent feature of all cells and represents DNA that contains no genes and so can always be retained in a compact organization. This fraction includes centromeric and telomeric DNA as well as certain regions of some other chromosomes. For example, most of the human Y chromosome is made of constitutive heterochromatin.

2. *Facultative heterochromatin* is not a permanent feature but is seen in some cells some of the time. Facultative heterochromatin is thought to contain genes that are inactive in some cells or at some periods of the cell cycle. When these genes are inactive, their DNA regions are compacted into heterochromatin.

It is assumed that the organization of heterochromatin is so compact that proteins involved in gene expression simply cannot access the DNA. In contrast, the remaining regions of chromosomal DNA, the parts that contain active genes, are less compact and permit entry of the expression proteins. These regions are called euchromatin and they are dispersed throughout the nucleus.

First, the degree of chromatin packaging displayed by a segment of a chromosome determines whether or not genes within that segment are expressed. Second, if a gene is accessible, then its transcription is influenced by the precise nature and positioning of the nucleosomes in the region where the transcription initiation complex will be assembled.

2.1.1 Regulation of Gene Expression by Histone modifications: Histone proteins can undergo various types of modification, the best studied of these being histone acetylation— the attachment of acetyl groups to lysine amino acids in the N-terminal regions of each of the core molecules. These N termini form tails that protrude from the nucleosome core octamer and their acetylation reduces the affinity of the histones for DNA and possibly also reduces the interaction between individual nucleosomes that leads to formation of the 30 nm chromatin fiber. The histones in heterochromatin are generally unacetylated whereas those in functional domains are acetylated, a clear indication that this type of modification is linked to DNA packaging.

Histone acetyltransferases (**HATs**) —the enzymes that add acetyl groups to histones . For example, one of the first HATs to be discovered, the *Tetrahymena* protein called p55, which was known to activate assembly of the transcription initiation complex.

One way in which silencing can be implemented is by removing acetyl groups from histone tails, and hence reversing the transcription-activating effects of the HATs. This is the role of the histone deacetylases (**HDACs**).

2.2.2. Gene Regulation by DNA methylation: Chromatin modification is not the only process that can bring about genome silencing. DNA methylation

can also repress gene activity. DNA methylation can repress gene activity. Initially this was thought to be quite distinct from histone modification, but now we are beginning to see links between the two activities. In eukaryotes, cytosine bases in chromosomal DNA molecules are sometimes changed to 5-methylcytosine by the addition of methyl groups by enzymes called DNA methyltransferases. Cytosine methylation is relatively rare in lower eukaryotes but in vertebrates up to 10% of the total number of cytosines in a genome is methylated, and in plants the figure can be as high as 30%. The methylation pattern is not random, instead being limited to the cytosine in some copies of the sequences 5′—CG—3′ and, in plants, 5′ CNG 3′.

Two types of methylation activity have been distinguished. The first is maintenance methylation which, following genome replication is responsible for adding methyl groups to the newly synthesized strand of DNA at positions opposite methylated sites on the parent strand. The maintenance activity therefore ensures that the two daughter DNA molecules retain the methylation pattern of the parent molecule.

The second activity is ***de novo* methylation**, which adds methyl groups at totally new positions and so changes the pattern of methylation in a localized region of the genome. It was originally thought that the first DNA methyltransferase, Dnmt1, was responsible for both types of methylation in mammalian cells. It was subsequently discovered that knockout mice that have an inactivated gene for Dnmt1 can still carry out *de novo* methylation. This led to the search for new enzymes and the eventual discovery of **$Dnmt_3a$ and $Dnmt_3b$,** which are now considered to be the main *de novo* methylases of mammals, with Dnmt1 primarily responsible for the maintenance activity.

Methylation results in repression of gene activity. This has been shown by experiments in which methylated or unmethylated genes have been introduced into cells by cloning and their expression levels measured: expression does not occur if the DNA sequence is methylated. The link with gene expression is also apparent when the methylation patterns in chromosomal DNAs are examined, these showing that active genes are located in unmethylated regions. For example, in humans, 40—50% of all genes are located close to CpG islands, with the methylation status of the CpG island reflecting the expression pattern of the adjacent gene. Housekeeping genes —those that are expressed in all tissues —have unmethylated CpG islands, whereas tissue-specific genes are unmethylated only in those tissues in which the adjacent gene is expressed.

How methylation influences genome expression was a puzzle for many years. Now it is known that methyl-CpG-binding proteins (**MeCPs**) are components of both the Sin3 and NuRD histone deacetylase complexes. This discovery has led to a model in which methylated CpG islands are the target sites for attachment of HDAC complexes that modify the surrounding chromatin in order to silence the adjacent genes.

Dosage Compensation by X chromosome inactivation in Mammals: It occurs because females have two X chromosomes whereas males have only one. If both of the female X chromosomes were active then proteins coded by genes on the X chromosome might be synthesized at twice the rate in females compared with males. To avoid this undesirable state of affairs, one of the female X chromosomes is silenced and is seen in the nucleus as a condensed structure called the Barr body, which is comprised entirely of heterochromatin. Silencing occurs early in embryo development and is controlled by the X inactivation center, a discrete region present on each X chromosome. In a cell undergoing X inactivation, the inactivation center on one of the X chromosomes initiates the formation of heterochromatin, which spreads out from the nucleation point until the entire chromosome is affected, with the exception of a few short segments containing small clusters of genes that remain active. The process takes several days to complete.

2.2. Regulatory Sequences and Transacting Factors:

'*Cis*' and '*trans*' are two important terms relevant to the genetic study of gene regulation in bacteria and eukaryotes.

- A locus is *cis*-acting on a second locus if it must be on the same DNA molecule in order to have an effect. The operator is a *cis*-acting element because it works only when physically attached to the gene whose expression it regulates.
- A locus is *trans*-acting if it can affect a second locus even when on a different DNA molecule. The gene for the lactose repressor (*lacI*) is *trans*-acting because it can regulate expression of the lactose operon even when removed from the *Escherichia coli* chromosome and placed on a plasmid.

To a molecular biologist, a *cis*-acting regulatory element is usually a target site for a DNA-binding protein, upstream of the gene whose expression is being regulated. A *trans*-acting element is the regulatory protein itself, which can diffuse through the cell from its site of synthesis to its DNA-binding site.

2.2.1 RNA Polymerase II Genes: Promoters and Enhancers (Regulatory Seqeunces):

RNA polymerase II (RNA Pol II) is located in the nucleoplasm. It is responsible for the transcription of all protein-coding genes and some small nuclear RNA genes. The pre-mRNAs must be processed after synthesis by cap formation at the 5'-end of the RNA and poly(A) addition at the 3'-end, as well as removal of introns by splicing.

Promoters: Many eukaryotic promoters contain a sequence called the **TATA box** around 25-35 bp upstream from the start site of transcription ***(Fig.* 1**). It has the 7 bp consensus sequence 5'-TATA(A/T)A(A/T)-3' although it is now known that the protein which binds to the TAT A box, TBP, binds to an 8 bp sequence that includes an additional downstream base pair, whose identity is not important. The TAT A box acts in a similar way to an E. *coli* promoter -10 sequence to position the RNA Pol II for correct transcription initiation. While the sequence around the TATA box is critical, the sequence between the TAT A box and the transcription start site is not critical. However, the spacing between the TATA box and the start site is important. Around 50% of the time, the start site of transcription is an adenine residue.

Some eukaryotic genes contain an initiator element instead of a TATA box. The initiator element is located around the transcription start site. Many initiator elements have a C at position -1 and an A at +1. Other promoters have neither a TAT A box nor an initiator element. These genes are generally transcribed at low rates, and initiation of transcription may occur at different start sites over a length of up to 200 bp. These genes often contain a GC-rich 20-50 bp region within the first 1-200 bp upstream from the start site.

Upstream regulatory elements: The low activity of basal promoters is greatly increased by the presence of other elements located upstream of the promoter. These elements are found in many I genes which vary widely in their levels of expression in different tissues. Two common examples are the **SP1** box, which is found upstream of many genes both with and without **TATA boxes**, and the **CCAAT box**. Promoters may have one or both or multiple copies of these sequences. These sequences which are often located within 1-200 bp upstream from the promoter are referred to as **upstream regulatory elements** (UREs) and play an important role in ensuring efficient transcription from the promoter.

Enhancers: Transcription from many eukaryotic promoters can be stimulated by control elements that are located many thousands of base pairs away from the transcription start site. This was first observed in the genome of the DNA virus SV40. A sequence of around 100 bp from SV40 DNA can significantly increase transcription from a basal promoter even when it is placed far upstream. Enhancer sequences are characteristically 10-200 bp long and contain multiple sequence elements which contribute to the total activity or the enhancer. They may be ubiquitous or cell type-specific. Classically, enhancers have the following general characteristics:

- They exert strong activation of transcription of a linked gene from the correct start site.
- They activate transcription when placed in either orientation with respect to linked genes.
- They are able to function over long distances of more than 1 kb whether from an upstream or downstream position relative to the start site.
- They exert preferential stimulation of the closest of two tandem promoters.

However, as more enhancers and promoters have been identified, it has been shown that the upstream promoter and enhancer motifs overlap physically and functionally. There seems to be a continuum between classic enhancer elements and those promoter elements which are orientation specific and must be placed close to the promoter to have an effect on transcriptional activity.

2.2.2. Regulation of Transcription Initiation

Transcription initiation is the stage at which the critical controls over the expression of individual genes (i.e. those controls that have greatest impact on the biochemical properties of the cell) are exerted. This is perfectly understandable. It makes sense that transcription initiation, being the first step in genome expression, should be the stage at which 'primary' regulation occurs, this being the level of regulation that determines which genes are expressed. we looked at how chromatin structure can influence gene expression by controlling the accessibility of promoter sequences to RNA polymerase and its associated proteins. This is just one way in which initiation of transcription can be regulated.

With bacteria, it is possible to make a clear distinction between constitutive and regulatory forms of control over transcription initiation. In bacteria, the RNA polymerase has a strong affinity for its promoter and the basal rate of transcription initiation is relatively high for all but the weakest promoters. With most eukaryotic genes, the reverse is true. The RNA polymerase II and III pre-initiation complexes do not assemble efficiently and the basal rate of transcription initiation is therefore very low, regardless of how 'strong' the promoter is. In order to achieve effective initiation, formation of the complex must be activated by additional proteins. Some of these could be defined as 'constitutive' activators, in that they work on many different genes and seem not to respond to any external signals; others could be termed 'regulatory' activators because they target a limited number of genes and do respond to external signals.

2.2.3 Constitutive transcription factor (SP1):

SP1 binds to a GC-rich sequence with the consensus sequence GGGCGG. It is a constitutive transcription factor whose binding site is found in the promoter of many housekeeping genes. SPI is present in all cell types. It contains three zinc finger motifs and has been shown to contain two glutamine-rich transac-tivation domains. The glutamine-rich domains of SPI have been shown to interact specifically with $TAF_{II}110$, one of the $TAF_{II}s$ which bind to the TATA -binding protein (TBP) to make up TFIID. This represents one target through which SPI may interact with and regulate the basal transcription complex.

2.2.4 Activators of eukaryotic transcription initiation

Any protein that stimulates transcription initiation is called an activator. Initially it was imagined that all activators were sequence-specific DNA-binding proteins, some recognizing upstream promoter elements and influencing transcription initiation only at the promoter to which these elements are attached, and others targeting sites within enhancers and influencing transcription of several genes at once. As with bacteria, eukaryotic enhancers can be some distance from their genes; their target specificity is ensured by the presence of insulators at either side of each functional domain, preventing the enhancers within that domain from influencing gene expression in adjacent domains. Whether bound to an upstream promoter element or to a more distant enhancer, the activator, according to the traditional view, stabilizes the pre-initiation complex by making contact with it.

This traditional view still holds for the majority of activators that have been identified but cannot be looked upon as all-encompassing. We have already seen that some proteins that were initially identified as activators are now recognized as components of chromatin modification complexes such as SAGA and Swi/Snf. Other proteins classed as activators influence gene expression by introducing bends and other distortions into DNA , possibly as a prelude to chromatin modification, or possibly to bring together proteins attached to non-adjacent sites, enabling the bound factors to work together in a structure that has been called an enhanceosome. An example of an activator that works in this way is SRY, which is the primary protein responsible for determining sex in mammals. Still other activators have no DNA-binding properties and they stimulate transcription simply by forming protein-protein contacts with the pre-initiation complex.

Activators have been looked upon as important in initiation by RNA polymerases II and III, but their role at RNA polymerase I promoters has been less well defined. RNA polymerase I is unusual in that it transcribes just a single set of genes: the multiple copies of the transcription unit containing the 28S, 5.8S and 18S rRNA sequences. These genes are expressed continuously in most cells, but the rate of transcription varies during the cell cycle and is subject to a certain amount of tissue-specific regulation. The regulatory mechanism has not been described in detail but recent research has suggested a role for the RNA polymerase I termination factor. This factor, called TTF-1 in mice and Reb1p in *Saccharomyces cerevisiae*, was first identified as an activator of RNA polymerase II transcription. It appears that the termination factor may also activate RNA polymerase I transcription, a binding site for it having been located immediately upstream of the promoter for the rRNA transcription unit.

Contacts between activators and the pre-initiation complex: A critical feature of the 'traditional' type of activator - those that bind to upstream promoter elements or to enhancers - is the contact that is formed with the pre-initiation complex. The part of the activator that makes this contact is called the activation domain. Structural studies have shown that although activation domains are variable, most of them fall into one of three categories:

- Acidic domains are ones that are relatively rich in acidic amino acids (aspartic acid and glutamic acid). This is the commonest category of activation domain.
- Glutamine-rich domains are often found in activators whose DNA-binding motifs are of the homeodomain or POU type.
- Proline-rich domains are less common.

A number of protein-protein interaction studies had suggested that direct contacts could be made between different activators and various parts of the complex, with TBP, various TAFs, TFIIB, TFIIH and RNA polymerase II all implicated as partners in different interactions. An alternative possibility was raised when a large protein complex called the mediator was identified in yeast. The mediator forms a physical contact between activators and the C-terminal domain of RNA polymerase II, suggesting that rather than direct interaction between an activator and the pre-initiation complex, the signal is transduced by the mediator. This hypothesis was strengthened when it was shown that the mediator possesses a protein kinase activity that enables it to phosphorylate the CTD of RNA polymerase II, stimulating promoter clearance. Current opinion tends to the view that a mediator is an obligatory component of the RNA polymerase II pre-initiation complex, and that the stimulatory effects of all activators pass through the mediator. The possibility that some activators bypass the mediator and have a direct effect on one or other part of the pre-initiation complex cannot, however, be discounted.

2.2.5 Repressors of eukaryotic transcription initiation

Most of the research on regulation of transcription initiation in eukaryotes has concentrated on activation, partly because the low level of basal initiation occurring at RNA polymerase II and III promoters suggests that the repression of initiation, which is so important in bacteria, is unlikely to play a major part in control of eukaryotic transcription. This view is probably incorrect because a growing number of DNA-binding proteins that repress transcription initiation are being discovered, these proteins binding to upstream promoter elements or to more distant sites in silencers. Some influence genome expression in a general way through histone deacetylation or DNA methylation, but others have more specific effects at individual promoters. The yeast repressors called Mot1 and NC2, for example, inhibit assembly of the pre-initiation complex by binding directly to TBP and disrupting its activity. Mot1 causes TBP to dissociate from the DNA, and

NC2 prevents further assembly of the complex on the bound TBP.

Another indication of the importance of repression in eukaryotic transcription comes from the demonstration that some proteins can exert both activating and repressing effects, depending on the circumstances. NC2, for example, represses initiation of transcription from promoters with a TATA box but has an activating effect on promoters that lack the TATA sequence.

Relatively little is known about the precise interactions occurring between repressors and the pre-initiation complex. A variety of inhibition domains (the converse of an activation domain) have been identified in eukaryotic repressors, several of which are rich in prolines, but no general patterns have emerged. The direct interactions with TBP displayed by Mot1 and NC2 argue against the involvement in repression of a complex equivalent to the mediator that is required for gene activation.

2.2.6 Hormonal regulation: Steroid Hormone Receptor:

Many transcription factors are activated by hormones which are secreted by one cell type and transmit a signal to a different cell type. One class of hormones, the steroid hormones, are lipid soluble and can diffuse through cell membranes to interact with transcription factors called steroid hormone receptors. In the absence of the steroid hormone, the receptor is bound to an inhibitor, and located in the cytoplasm *(Fig.* 1).

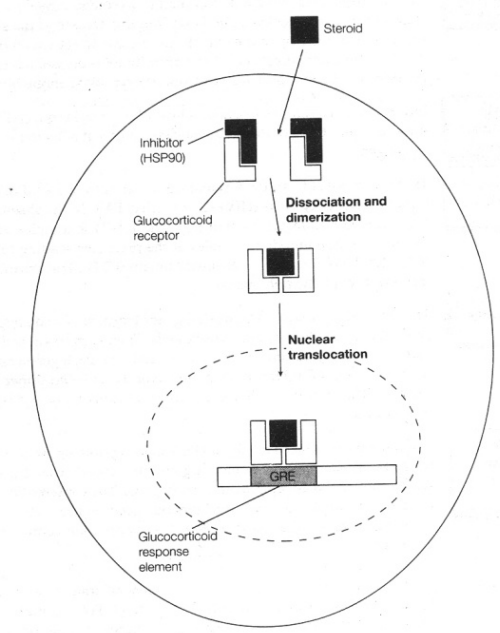


Fig. 1. Steroid hormone activation of the glucocorticoid receptor.

The steroid hormone binds to the receptor and releases the receptor from the inhibitor, allowing the receptor to dimerize and translocate to the nucleus. The DNA-binding domain of the steroid hormone receptor then interacts with its specific DNA-binding sequence, or **response element**, and this gives rise to activation of the target gene. Important classes of related receptors include the glucocorticoid, estrogen, retinoic acid and thyroid hormone receptors. The general model described above is not true for all of these. For example, the thyroid hormone receptors act as DNA-bound repressors in the absence of hormone. In the presence of the hormone, the receptor is converted from a transcriptional repressor to a transcriptional activator.

2.2.7 Regulation by Phosphorylation: STAT proteins:

Many hormones do not diffuse into the cell. Instead, they bind to cell-surface receptors and pass a signal to proteins within the cell through a process called signal transduction. This process often involves protein phosphorylation. Interferon-γ induces phosphorylation of a transcription factor called STATα through activation of the intracellular kinase called Janus activated kinase (JAK). When STAT 1α protein is unphosphorylated, it exists as a monomer in the cell cytoplasm and has no transcriptional activity. However, when STAT1α becomes phosphorylated at a specific tyrosine residue, it is able to form a homodimer which moves from the cytoplasm into the nucleus. In the nucleus, STAT1α is able to activate the expression of target genes whose promoter regions contain a consensus DNA-binding motif.

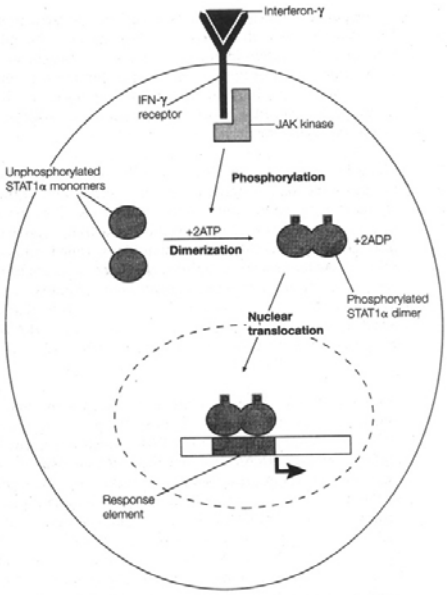


Fig. 2. Interferon-γ-mediated transcription activation caused by phosphorylation and dimerization of the STAT1α transcription factor.

2.2.8 Embryonic Development: The homeobox is a conserved DNA sequence which encodes the helix-turnhelix DNA binding protein structure called the homeodomain. The homeodomain was first discovered in the transcription factors encoded by homeotic genes of *Drosophila.* Homeotic genes are responsible for the correct specification of body parts. For example, mutation of one of these genes,

Antennapedia, causes the fly to form a leg where the antenna should be. These genes are very important in spatial pattern formation in the embryo. The homeobox sequence has been conserved between a wide range of eukaryotes and homeobox-containing genes have been shown to be important in mammalian development. In *Drosophila* and mammals, the homeobox genes are arranged in gene clusters in which homologous genes are in the same order. The gene homologs are also expressed in a similar order in the embryo on the anterior to posterior axis. This suggests that the conserved homeobox-encoded DNA-binding domain is characteristic of transcription factors which have a conserved function in embryonic development.

2.2.9 Translational Control

Because of the different natures of the mRNA in prokaryotes and eukaryotes (i.e. polycistronic VS. monocistronic) and the absence of the nuclear membrane in the former, different possibilities exist for the control of translation. In prokaryotes, the structure formed by regions of the mRNA can obscure ribosome binding sites, thus reducing translation of some cistrons relative to others. The formation of stems and loops can inhibit exonucleases and give certain regions of the polycistronic mRNA a greater half-life (and hence a greater chance of translation) than others. Several operons encoding ribosomal proteins show an interesting form of translational control in E. *coli* where a region of the mRNA has a tertiary structure that resembles the binding site for a ribosomal protein encoded by the mRNA. If there is insufficient rRNA available for the translation product to bind to, it will bind to its own mRNA and prevent further translation. Prokaryotes sometimes make short **antisense RNA** molecules that form duplexes near the ribosome binding site of certain mRNAs, thus inhibiting translation.

Eukaryotes generally control the amount of specific proteins by varying the level of transcription of the gene and/or by RNA processing, but some controls occur in the cytoplasm. The presence of multiple copies of **5'-AUUUA-3'**, usually in the 3'-noncoding region, marks the mRNA for rapid degradation and thus limited translation. Another form of translational control involves proteins binding directly to the mRNA and preventing translation. This RNA is called '**masked mRNA**'. In appropriate circumstances, the mRNA can be translated when the protein dissociates. Some noncoding sequences can cause mRNA to be located in a specific part of the cytoplasm and, when translated, can give rise to a gradient of protein concentration across the cell.

Poly proteins: Bacteriophage and viral transcripts and many mRNAs for hormones in eukaryotes (e.g. pro-opiomelanocortin) are translated to give a single polypeptide chain that is cleaved subsequently by specific proteases to produce multiple mature proteins from one translation product. The parent polypeptide is called a **poly protein**.

Protein Targeting: It has been discovered that the ultimate cellular location of proteins is often determined by specific, relatively short, amino acid sequences within the proteins themselves. These sequences can be responsible for proteins being secreted, imported into the nucleus or targeted to other organelles. The greater complexity of the eukaryotic cell means that there are more types of targeting in eukaryotes. Protein secretion in both prokaryotes and eukaryotes involves a signal sequence in the nascent protein and specific proteins or, in the latter, an RNP particle, **signal recognition particle (SRP)**, that recognizes it *(Fig.1).*

If a cytosolic ribosome begins to translate an mRNA encoding a protein that is to be secreted, SRP binds to the ribosome and the emerging polypeptide and arrests translation. SRP is capable of recognizing ribosomes with a nascent chain containing a signal sequence (signal peptide) which is composed of about 13-36 amino acids having at least one positively charged residue followed by a hydrophobic core of 10-15 residues followed by a small, neutral residue, often Ala. SRP with the arrested ribosome binds to a receptor (SRP receptor or docking protein) on the cytosolic side of the endoplasmic reticulum (ER) and, when the ribosome becomes attached to ribosome receptor proteins on the ER, SRP is released and can be re-used. The ribosome is able to continue translation, and the nascent polypeptide chain is pushed through into the lumen of the ER. As it passes through, signal peptidase cleaves off the signal peptide. When the protein is released into the ER it is usually modified, often by glycosylation, and different patterns of glycosylation seem to control the final location of the protein.

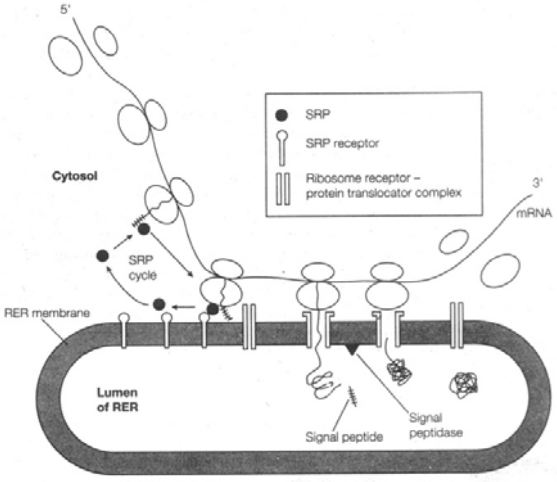


Fig. 1. Protein secretion in eukaryotes.

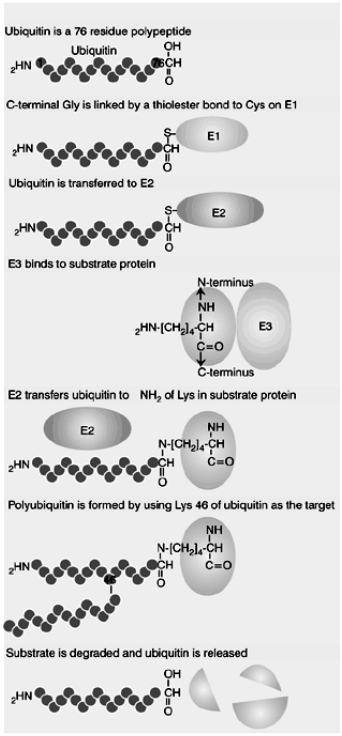
Other peptide sequences in proteins are responsible for their cellular location. Different N-terminal sequences can cause proteins to be imported into mitochondria or chloroplasts, and the internal sequence -Lys-Lys-Lys-Arg-Lys, or any five consecutive positive amino acids, can be a nuclear localization signal (NLS) causing the protein containing it (e.g. histone) to be imported into the nucleus.

Protein Modification: A newly translated polypeptide does not always immediately generate a functional protein. Apart from correct folding and the possible formation of disulfide bonds, there are a number of other alterations that may be required for activity. These include cleavage and various covalent modifications. Cleavage is very common, especially trimming by amino- and carboxypeptidases, but the removal of internal peptides also occurs, as in the case of insulin. Signal sequences are also usually cleaved off secreted proteins and, where proteins are made as parts of polyproteins, they must be cleaved to release the component proteins. Ubiquitin is made as a polyprotein containing multiple copies linked end-to-end, and this must be cleaved to generate the individual ubiquitin molecules.

Chemical modifications are many and varied and have been shown to take place on the N and C termini, as well as on most of the 20 amino acid side chains, with the exception of Ala, Gly, lIe, Leu, Met and Val. The modifications include a,cetylation, hydroxylation, phosphorylation, methylation, glycosylation and even the addition of nucleotides. Hydroxylation of Pro is common in collagen, and some of the histone proteins are often acetylated. The activity of many enzymes, such as glycogen phosphorylase and some transcription factors, is controlled by phosphorylation.

Protein Degradation: Different proteins have very different half-lives. Regulatory proteins tend to turn over rapidly and cells must be able to dispose of faulty and damaged proteins. In eukaryotes, it has been discovered that the N-terminal residue plays a critical role in inherent stability. Eight N-terminal amino acids (Ala, Cys, Gly, Met, Pro, Ser, Thr, Val) correlate with stability ($t_{1/2}$ > 20 hours), eight.(Arg, His, lIe, Leu, Lys, Phe, Trp, Tyr) with short $t_{1/2}$ (2-30 min) and four (Asn, Asp, GIn, Glu) are destabilizing following chemical modification. A protein that is damaged, modified or has an inherently destabilizing N-terminal residue becomes ubiquitinylated by the covalent linkage of molecules of the small, highly conserved protein, ubiquitin, via its C-terminal Gly, to lysine residues in the protein. The ubiquitinylated protein is digested by a 26 S protease complex (proteasome) in a reaction that requires A TP and releases intact ubiquitin for re-use.

Figure 8.43 The ubiquitin cycle involves three activities. E1 is linked to ubiquitin. E3 binds to the substrate protein. E2 transfers ubiquitin from E1 to the substrate. Further cycles generate polyubiquitin.



D3. EUKARYOTIC TRANSCRIPTION FACTORS:

Transcription factors other than the general transcription factors of the basal transcription complex were first identified through their affinity for specific motifs in promoters, upstream regulatory elements (UREs) or enhancer regions. These factors have two distinct activities. Firstly, they bind specifically to their DNA-binding site and, secondly, they activate transcription. These activities can be assigned to separate protein domains called activation domains and DNA-binding domains. In addition, many transcription factors occur as homo- or heterodimers, held together by dimerization domains. A few transcription factors have ligand-binding domains which allow regulation of transcription factor activity by binding of an accessory small molecule. The steroid hormone receptors are an example containing all four of these types of domain.

Mutagenesis of the yeast transcription factors Gal4 and Gcn4 showed that their DNA-binding and transcription activation domains were in separate parts of the proteins. Experimentally, these activation domains were fused to the bacterial LexA repressor. These hybrid fusion proteins activated transcription from a promoter containing the *lexA* operator sequence, indicating that the transcriptional activation function of the yeast proteins was separable from their DNA-binding activity. These type of experiments are called domain swap experiments.

3.1 The helix-turn-helix domain

This domain is characteristic of DNA-binding proteins containing a 60-aminoacid homeodomain which is encoded by a sequence called the homeobox. In the Antennapedia transcription factor of *Drosophila,* this domain consists of four α-helices in which helices II and III are at right angles to each other and are separated by a characteristic β-tum.

The characteristic helix-turnhelix structure *(Fig.1*) is also found in bacteriophage DNA-binding proteins such as the phage A cro repressor, lac and trp

repressors, and cAMP receptor protein, CRP. The domain binds so that one helix, known as the recognition helix, lies partly in the major groove and interacts with the DNA. The recognition helices of two homeodomain factors Bicoid and Antennapedia can be exchanged, and this swaps their DNA-binding specificities. Indeed, the specificity of this interaction is demonstrated by the observation that the exchange of only one amino acid residue swaps the DNA-binding specificities.

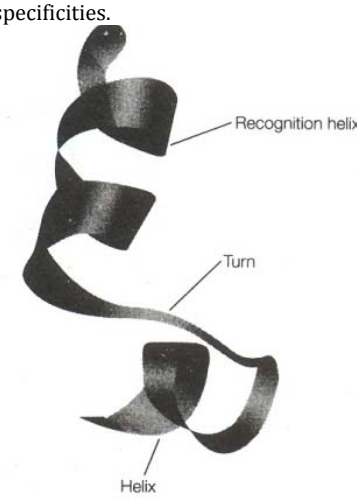


Fig. 1. The helix–turn–helix core structur

3.2 The zinc finger domain:

This domain exists in two forms. The C_2H_2 zinc finger has a loop of 12 amino acids anchored by two cysteine and two histidine residues that tetrahedrally co-ordinate a zinc ion *(Fig.2).* This motif folds into a compact structure comprising two β-strands and one α-helix, the latter binding in the major groove of DNA *(Fig. 2b)* The a-helical region contains conserved basic amino acids which are responsible for interacting with the DNA. This structure is repeated nine times in TFIIIA, the RNA Pol III transcription factor. It is also present in transcription factor SPI (three copies). Usually, three or more C_2H_2 zinc fingers are required for DNA binding. A related motif, in which the zinc ion is co-ordinated by four cysteine residues, occurs in over 100 steroid hormone receptor transcription factors. These factors consist of homo- or hetero-dimers, in which each monomer contains two C_4 '**zinc finger**' motifs *(Fig.).*

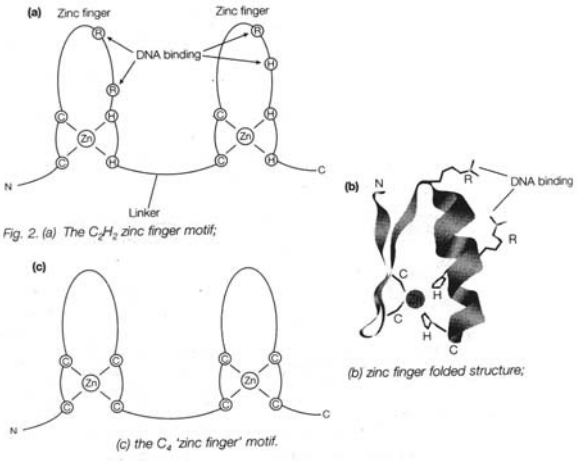


Fig. 2. (a) The C_2H_2 zinc finger motif;

(b) zinc finger folded structure;

(c) the C_4 'zinc finger' motif.

The two motifs are now known to fold together into a more complex conformation stabilized by zinc, which binds to DNA by the insertion of one a-helix from each monomer into successive major grooves, in a manner reminiscent of the helix-turn-helix proteins.

The basic domain: A basic domain is found in a number of DNA-binding proteins and is generally associated with one or other of two dimerization domains, the leucine zipper or the helix-loop-helix (HLH) motif. These are referred to as basic leucine zipper (bZIP) or basic HLH proteins. Dimerization of the proteins brings together two basic domains which can then interact with DNA.

3.3. Leucine zippers

Leucine zipper proteins contain a hydrophobic leucine residue at every seventh position in a region that is often at the C-terminal part of the DNA-binding domain. These leucines lie in an α-helical region and the regular repeat of these residues forms a hydrophobic surface on one side of the α-helix with a leucine every second turn of the helix. These leucines are responsible for dimerization through interactions between the hydrophobic faces of the α-helices (see *Fig*). This interaction forms a coiled-coil structure. bZIP transcription factor contain a basic DNA-binding domain N-terminal to the leucine zipper. This is present on an α-helix which is a continuation from the leucine zipper α-helical C-terminal domain. The N-terminal basic domains of each helix form a symmeterical structure in which each basic domain lies along the DNA in opposite directions, interacting with a symmetrical DNA recognition site so that the protein in effect forms a clamp around the DNA. The leucine zipper is also used as a dimerization domain in proteins that use DNA-binding domains other than the basic domain, including some homeodomain proteins.

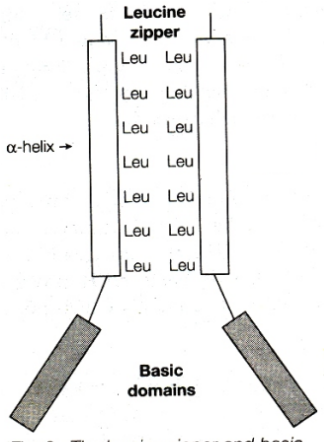


Fig. 3. The leucine zipper and basic domain dimer of a bZIP protein.

The helix-loop-helix domain: The overall structure of this domain is similar to the leucine zipper, except that a non helical loop of polypeptide chain separates two α-helices in each monomeric protein. Hydrophobic residues on one side of the C-terminal a-helix allow dimerization. This structure is found in the **MyoD** family of proteins. As with the leucine zipper, the HLH motif is often found adjacent to a basic domain that requires dimerization for DNA binding. With both basic HLH proteins and bZIP proteins the formation of heterodimers allows much greater diversity and complexity in the transcription factor repertoire.

3.4 Transcription Activation Domain:

Acidic activation domains: Comparison of the transactivation domains of yeast Gcn4 and Ga14, mammalian glucocorticoid receptor and herpes virus activator VPI6 shows that they have a very high proportion of acidic amino acids. These have been called acidic activation domains or 'acid blobs' or 'negative noodles' and are characteristic of many transcription activation domains. It is still uncertain what other features are required for these regions to function as efficient transcription activation domains.

Glutamine-rich domains: Glutamine-rich domains were first identified in two activation regions of the transcription factor SPI. As with acidic domains, the proportion of glutamine residues seems to be more important than overall structure. Domain swap experiments between glutamine-rich transactivation regions from the diverse transcription factors SPI and the *Drosophila* protein Antennapedia showed that these domains could substitute for each other.

Proline-rich domains: Proline-rich domains have been identified in several transcription factors. As with glutamine, a continuous run of proline residues can *activate* transcription. This domain is found, for example, in the c-Jun, AP2 and Oct-2 transcription factors.

3.5 Repressor Domain:

Repression of transcription may occur by indirect interference with the function of an activator. This may occur by:

- Blocking the activator DNA-binding site (as with prokaryotic repressors;
- Formation of a non-DNA-binding complex (e.g. the repressors of steroid hormone receptors, or the Id protein which blocks HLH protein-DNA interactions, since it lacks a DNA-binding domain;.
- Masking of the activation domain without preventing DNA binding (e.g. *Gal80* masks the activation domain of the yeast transcription factor Gal4).

In other cases, a specific domain of the repressor is directly responsible for inhibition of transcription. For example, a domain of the mammalian thyroid hormone receptor can repress transcription in the absence of thyroid hormone and activates transcription when bound to its ligand. The product of the Wilms tumor gene, *WT1* is a tumor-suppressor protein having a specific proline-rich repressor domain that lacks charged residues.

3.6 Targets of Transcriptional Regulation:

The presence of diverse activation domains raises the question of whether they each have the same target in the basal transcription complex or different targets for the activation of transcription. They are distinguishable from each other since the acidic activation domain can activate transcription from a downstream enhancer site while the proline domain only activates weakly and the glutamine domain not at all. While proline and acidic domains are active in yeast. glutamine domains have no activity, implying that they have a different transcription target which is not present in the yeast transcription complex. Proposed targets of different transcriptional activators include:

- chromatin structure;
- interaction with TFIID through specific $TAF_{II}s$;
- interaction with TFIIB;
- interaction or modulation of the TFIIH complex activity leading to differential phosphorylation of the CTD of RNA Pol II.

lt seems likely that different activation domains may have different targets, and almost any component or stage in initiation and transcription elongation could be a target for regulation resulting in multistage regulation of transcription.

IFAS

JODHPUR